

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

A Critical Exploration of Philip Pettit's Theory of Group Agency

By Leo Townsend

**A minor dissertation submitted in partial fulfilment of the requirements for the award of
the degree of Master of Arts in Philosophy**

Faculty of the Humanities

University of Cape Town

2012

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: _____ Date: _____

TABLE OF CONTENTS

Abstract.....	3
Acknowledgement.....	5
Introduction: Committees and Communities.....	6
Essay 1: Aggregation and Alienation: from Lilliputians to Leviathans	13
Essay 2: Doctrine and Reason.....	35
Essay 3: Being and Becoming in the Theory of Group Agency.....	54
Afterword: Whither Group Agency?	74
Bibliography.....	76

Abstract

There has recently been renewed interest within analytic philosophy in the nature and status of social collectives or groups. A number of contemporary theorists have claimed that collectives share certain characteristics with the individuals who comprise them – that groups, just like individuals, have beliefs and desires; that they speak and act; that they should be held responsible for their actions and utterances; and that they should be accorded certain rights and entitlements.

Philip Pettit's 'theory of group agency' is an interesting example of this recent development. While giving methodological and normative primacy to individuals, he nonetheless holds that certain groups should be recognised as rational agents in their own rights, as corporate persons with minds of their own. Pettit's support for these claims derives from three primary sources. First, he endorses a broadly 'functionalist' approach to mindedness and agency, according to which being an intentional subject or agent is a matter of being able to perform in a certain way. Second, he draws on results obtained in the field of 'judgement aggregation', which suggest that the collectively-held attitudes of a group cannot be a straightforward function of the attitudes of its members. Third, he employs some of the theoretical resources recently developed in the literature on 'collective intentionality', especially the notions of joint action and shared intention.

This dissertation is a critical response to Pettit's theory. It consists of three self-standing essays, each of which targets one of sources of support mentioned above.

Essay 1 is concerned with Pettit's claim that groups can count as corporate persons. It initially draws a parallel between Pettit's theory and a theory of mind called 'homuncular functionalism'. Both theories portray persons as corporate entities or 'integrates', and this appears to secure a welcome autonomy for persons from their underpinnings. In the latter part of this essay, however, I pursue a line of objection against this corporate approach to

personhood, arguing that it relies upon an understanding of intentionality that is radically alienating.

Essay 2 focuses on Pettit's influential interpretation of the 'Doctrinal Paradox', a problem recently observed to affect legal adjudication in multi-member courts. Pettit's claim, which forms the cornerstone of his theory of group agency, is that this problem is far more general and far more serious than was first imagined. I dispute this, arguing that none of Pettit's explicit generalising moves is successful.

Essay 3 argues that Pettit's account of the genesis of group agents faces a particular sort of bootstrapping problem that he has not successfully resolved. Pettit claims that group agents emerge when a collective jointly acts – on the basis of their shared intention – to exercise reason at the collective level, but I question whether this is something a collective can accomplish without already being a group agent.

Each essay also supplements its negative case against Pettit by gesturing at a positive communitarian alternative inspired by the work of Robert Brandom. My hope is that this dissertation will be a prolegomenon to a future fuller development of this alternative.

Acknowledgement

I wish to express my gratitude to a number of people who have helped me in various ways to complete this dissertation. My supervisor, Dr Greg Fried, deserves special thanks for sparking my interest in this topic and guiding me through the writing process with patience and commitment. I have benefited enormously from his comments on draft versions of this text, as well as from feedback provided by Elizabeth Braae and Prof. Bernhard Weiss. I would also like to thank my family and friends for their constant support, and my former supervisor, Dr Jeremy Wanderer, for his ongoing encouragement of my philosophical aspirations.

I dedicate this dissertation to my wife, Dina Townsend, who made it possible, and to my daughter, Jemima Townsend, who made it doable.

Introduction: Committees and Communities

1. Context

Corporate personhood has lately emerged as a topic of intense debate, both for philosophers and the public at large. Much of the recent popular interest arose in response to the US Supreme Court's controversial ruling in *Citizens United vs FEC*,¹ which granted corporations the right to engage in overt political expenditure on the basis of their constitutionally enshrined personhood. For many this decision seemed to confer upon corporations rights which should be reserved for individual human beings, sparking dystopian visions of a future where human interests are superseded by those of commercial conglomerations. If this populist backlash to the Supreme Court's ruling had a unifying conviction it was that corporations are 'persons' only so far as it is beneficial for humans to treat them as such – whereas the personhood of humans is intrinsic and inalienable.

Within analytic philosophy, however, a contrasting reversal has been steadily gathering steam. In the philosophy of mind and personal identity, the rise of functionalism has seen many philosophers dismiss as 'chauvinism' (Block 1980) the idea that personhood is essentially linked to human-ness. The personhood or mindedness of a creature, they contend, depends not on its intrinsic features but on how the creature is able to perform, and so, in principle, the conditions of personhood could be met by social creatures – institutions, firms, perhaps even whole nations.²

Just as humans are being dislodged from positions of privilege in theorising and minds and persons, so too, within analytic social philosophy, is the centrality of the individual being called into question. Mid-twentieth century social philosophy was dominated by the 'methodological individualism' of Popper and Hayek, the "important doctrine that all social phenomena... should always be understood as resulting from the decisions, actions, attitudes, etc., of human individuals, and that we should never be satisfied by an explanation in terms of so-called 'collectives'" (Popper 1945: 91). In terms that foreshadow today's populist backlash, the propagators of this orthodoxy linked the rival 'collectivist' mode of social explanation to

¹ 558 U.S., 130 S. Ct. 876 (2010)

² See, e.g., Brooks (1986), Biro (1981), Lycan (1987), Dennett (1978), Block (1980), Huebner et al (2010)

the perils of totalitarianism,³ and dismissed as ‘unscientific’ the collective psychological forces it appeared to invoke. But in recent times there has been a renewed interest within mainstream analytic philosophy in a cluster of concerns broadly related to *the collective*, including joint action, collective intentionality, social ontology and, most recently, group agency. Today’s analytic philosophers are increasingly unashamed to advance claims about ‘shared agency’ (Bratman 2009), distinctive ‘we-modes’ (Tuomela 2003) of intention, institutional facts (Searle 1995), ‘plural subjects’ (Gilbert 1989) and ‘group persons’ (Rovane 1997).

Philip Pettit’s theory of group agency⁴ – the position I will be critically exploring in detail in the three essays which follow – is, in one sense, part of this recent philosophical movement. Endorsing a broadly functionalist approach to agency, personhood and mind, Pettit argues that some social groups should be counted as ‘rational agents in their own rights’ (List & Pettit 2011) – as ‘corporate persons,’ (Pettit 2002) with ‘minds of their own’ (Pettit 2003). But in advancing these claims Pettit will have no truck with the collectivist thinking that has lately increased its presence in analytic philosophy.⁵ In fact, Pettit’s postulation of group agents is to some extent *premised* upon his adherence to methodological individualism, and the curtailments he recommends to the rights and freedoms of corporate persons derive from his commitment to ‘normative individualism’ – the view that the only true bearers of value are individuals.

Pettit submits that his theory is *historically novel*, since it departs from both of the traditional realist perspectives on group agency. The first of these is the ‘authorisation theory’ associated

³ See, e.g., Popper (1962, Ch. 12)

⁴ See esp. Pettit (2001, 2002, 2003, 2007, 2009); List & Pettit (2006, 2011); Pettit & Schweikard (2006). Throughout this dissertation I refer to the theory of group agency as Pettit’s theory, even though it has been largely worked out in collaborative publications (most regularly with Christian List). This is done for the sake of convenience and is certainly not meant to undervalue the contributions of his collaborators. Nonetheless, Pettit is the common thread in all the publications I am focused on, and, judging by his individually-authored work, it seems likely that he is responsible for the sections of the co-authored work of primary interest to me here.

⁵ One question I do not tackle in this dissertation is how Pettit’s thinking over the past ten years relates to the ‘social holism’ he advanced in *The Common Mind* (1993). Social holism was there defended as the view that the ability of certain creatures to *think* is both causally and non-causally dependent on their social situation. Though I do not think Pettit has changed his mind, it is striking to find that *Group Agency* (2011) – in many respects the culmination of his past decade’s work – contains no mention of this thesis which figured so prominently in earlier work. So far as I know Pettit has not explicitly sought to synthesize these two phases of his writing, and neither have I taken on that task here.

with Hobbes, according to which a group agent is formed when each member of a multitude authorises a sub-group or individual to speak and act on their behalf. Pettit contends that this realism about group agents is ‘redundant’, since “everything the recognition of such a group agent entails is already expressible in individual-level language” (List & Pettit 2011: 8). However, Pettit also wishes to distance his theory from the other venerable tradition of theorising about group agents, which he calls ‘animation theory’. Animation theory, which Pettit traces to the influence of Hegel, holds that group agents are formed when a transcendent force “animates them and gives rise to a single centre of agency” (List & Pettit 2011: 9). But this “offends against methodological individualism... [and] is metaphysically incredible” (List & Pettit 2011:9) and so Pettit dismisses this too.

Pettit’s theory thus seeks to avoid both the redundancy of the broadly Hobbesian approach and the mysteriousness of the Hegel-inspired alternative. He thinks he can develop a robust realism about group agents that is compatible with – indeed partly dependent on – an unwavering commitment to individualism. In framing his theory in these terms, Pettit draws frequent parallels between social ontology and the philosophy of mind, depicting the ‘metaphysically incredible’ Hegelian approach as the analogue of Cartesian dualism, and the ‘redundant’ Hobbesian position as corresponding to crude reductionism about minds (e.g., mind-brain identity theory). Crucially, Pettit construes the *individualism* underlying his account as playing a role similar to that played by the doctrine of *physicalism* in much recent philosophy of mind:⁶

“Many physicalists recognise that there are... minds, and consider this recognition important for our understanding of the world, but deny that this requires any physically mysterious stuff or force – anything like Descartes’s *res cogitans*.... Individualists can similarly hold that there are group agents, and that this recognition is important for a proper understanding of the social world, and yet deny that this requires any psychologically mysterious element” (List & Pettit 2011:6)

⁶ Though ‘physicalism’ is sometimes used more specifically to refer to type-identity or reductionist theories of mind, it is meant here in the sense of ‘minimal’ or ‘supervenience’ physicalism (Lewis 1983) – the quite general (though still contested) claim that everything is ultimately physical, and hence that mental states and events must supervene on physical states and events.

Much of Pettit's support for his theory of group agency can be seen as an attempt to make good on this analogy. In the end he wishes to present a *non-reductive individualism* about group agents that mirrors *non-reductive physicalism* about minds, and in this I think he is more or less successful. But this assessment is not the conclusion so much as the point of departure for my critical project, for I think that individualism in social ontology, just like physicalism in the philosophy of mind and action, is not a commendable commitment but a debilitating prejudice.⁷ Accordingly, the three essays which follow aim to describe some of the damaging effects of this prejudice, and to see how these might be avoided.

Another, subsidiary aim is to bring the work of Pettit into contact with that of another towering figure in analytic philosophy, Robert Brandom. Though both Pettit and Brandom have sought to emphasize the connections between mindedness and sociality, they have done this markedly different ways, and as a consequence the relevance of their views to one another has seldom been explored.⁸ This is, I think, a pity. In particular, I think that though Brandom has not himself made its social ontological consequences fully explicit, his Hegel-inspired inter-subjectivism is a promising antidote to the individualism which ultimately cripples Pettit's project. In what follows I gesture at this promise by appending a broadly Brandomian positive proposal to each of my essay's critique of Pettit. As I indicate in a brief afterword, I hope in future research to expand on these sketches, developing a broadly Brandomian alternative to Pettit's individualistic social ontology.

The contrast between the views of Pettit and Brandom which I try to exploit in these essays is perhaps most clearly visible by reflecting on what each chooses as his exemplar of sociality. For Pettit the salient sort of social entity is the *committee*. In a committee, a number of pre-existing individuals come together for the purpose of generating judgements or decisions about how to act, usually in line with some pre-agreed protocols and for the fulfilment of some pre-established mandate. Committees are centres of co-ordinated rational deliberation between separate rational agents; they involve sets of individuals thinking and acting

⁷ Towards the end of Essay 1 I suggest that individualism and physicalism may be varieties of the same prejudice, but apart from that my focus is on individualism and not physicalism.

⁸ Engel (2002) compares Brandom's (1994) views on discursive norms with the approach to intentionality developed by Pettit (1993). My aim, by contrast, is to consider the most recent developments in Pettit's thinking – i.e., the theory of group agency – in the light of Brandom's newer, more explicitly Hegelian work.

together, towards some common objective. What Pettit's distinctive approach to social ontology attempts to show us is that committee-like social entities often can and must take on a life of their own, becoming rational agents in their own rights, yet without being anything besides the individuals which underpin them.

Brandom's focus, by contrast, is on *communities*. Like a committee, a community is comprised of separate rational agents or persons, but Brandom's key message is that the status of these 'individuals' *as rational* or *as persons* is not intelligible apart from their membership and participation in the community. A community is a special sort of social entity then because it institutes and enforces certain normative statuses which are otherwise susceptible to being misconceived or obscured. Accordingly, a Brandom-inspired social ontology would be more concerned to emphasize the way personhood is a distinctively social status than with the possible status of social groups as persons. This is what I think Pettit's individualistic social ontology misses out on, and the burden of these essays is to bring out some of the consequences of this failing.

2. Essay Outlines

Though they share this overarching aim, each essay pursues a different *sort of question* about group agency: first a 'what-question' about *meaning*; second a 'why-question' about *motivations*; and third a 'how-question' about *mechanisms*.

Essay 1, 'Aggregation and Alienation: from Lilliputians to Leviathans', is focused on what Pettit's claim that groups can count as corporate persons or group agents *means*. In fact, the essay addresses two distinct *what-questions* about the meaning of these claims. First, in what *sense* might groups count as intentional subjects or persons? I answer this by drawing a parallel between Pettit's theory of group agency and a position in the philosophy of mind known as 'homuncular functionalism'. Both theories seek to portray persons as corporate entities or 'integrates', and this appears to secure a welcome autonomy for persons from their underpinnings.

The second *what-question* in this first essay asks after the conceptual implications of Pettit's theory: what *significance* does Pettit's recognition of groups as persons have for our concept

of personhood? My claim here is that the view of personhood which licenses Pettit's recognition of social groups as persons also ensures a kind of radical alienation, such that no genuine persons are ever on view. I end off with an alternative communitarian proposal inspired by Brandom, which sees personhood as a distinctively social achievement, with distinctively social implications. This alternative resuscitates a notion of genuine persons without lapsing into the 'chauvinism' characteristic of intrinsicist approaches to personhood.

Essay 2, 'Doctrine and Reason', addresses a why-question about the *motivations* of group agents: why should a collective seek to co-ordinate their attitudes and activity so as to constitute a single centre of rational agency? Answering this question gets us to attend to Pettit's influential interpretation of a problem recently observed to afflict legal adjudication in multi-member courts. The problem, in short, is that the individual rationality of judges – conceived in terms of their individual adherence to 'legal doctrine' – cannot guarantee the collective rationality of the court's group judgements (judgements derived, for instance, by majority voting). Pettit's claim, which forms the cornerstone of his theory of group agency, is that this judicial difficulty exemplifies a quite general 'Discursive Dilemma' which many kinds of groups, in many kinds of contexts, must inevitably confront.

This second essay aims to cast doubt upon Pettit's generalising move from the Doctrinal Paradox to the Discursive Dilemma. Pettit insists that the paradox can arise even when nothing plays the role played by 'doctrine' in the original jurisprudential examples, and my suggestion is that this is true only if reason itself is conceived in doctrinal terms. I conclude this essay by venturing a deflationary response to the Doctrinal Paradox inspired by Brandom's various discussions of legal rationality and the content of legal concepts. Like Pettit, Brandom takes legal adjudication to exemplify certain characteristics of discursive practice more generally, but Brandom's emphasis on the essentially communal character of discursive practice gives us pause to reconsider the force of both the Doctrinal Paradox and the Discursive Dilemma.

Essay 3, 'Being and Becoming in the Theory of Group Agency', is concerned with a *how-question* about the mechanisms of group agency. Assuming that certain sets of individuals have strong motivation for incorporating into group agents or persons, just *how* should they

go about doing this? Pettit's answer is that such collectives must strive to exercise reason at the collective level, and he describes several decision-making procedures which groups can adopt in order to achieve this.

My main claim in this last essay is that the procedures which purportedly give rise to corporate persons could only plausibly be adopted by something that was already a unified rational agent or person, and so it appears that no coherent account of the genesis of corporate persons is ever given by Pettit. I trace the problem to Pettit's commitment to individualism: for an individualist, group agents can only avoid redundancy if they can somehow bootstrap themselves into existence. This is a tremendously hard thing to explain for an individualist, but perhaps not so hard for a neo-Hegelian like Brandom. I conclude by suggesting that the personhood of groups might depend not only on the efforts of group members but also on the recognitive attitudes of other persons in a wider discursive community.

As these brief synopses suggest, each of the three essays has its own points of primary focus and its own ambitions. Each has also been written to be more or less self-standing – to be, as it were, an individual. But taken together, I hope that the three essays can mimic the unity of purpose and collective coherence that Pettit thinks is possible for groups. Though their aims and concerns sometimes diverge, the result should be variety rather than disharmony, for they are meant as discrete elements in a single critical exercise.

Essay 1: Aggregation and Alienation: from Lilliputians to Leviathans

Introduction

A certain, broadly functionalist or ‘performative’ (List & Pettit 2011: 171), approach to personhood allows us to recognise many more persons than those of human size and shape. To be a person is to be able to perform as a person – to ‘personate’, as Hobbes put it (Leviathan, Ch. 16) – and while such performance might require certain underlying states and mechanisms, these too can be functionally defined. This approach has the well-publicised ‘liberal’ consequence that persons can be *multiply realised*, and hence that creatures of non-human physiology or artificial provenance, and even those who lack subjective experience entirely, are not excluded from the community of persons on ‘chauvinist’ grounds (Block 1980). But though the prospect of recognising Martians, chess-playing computers or zombies as persons has dominated much of the debate around the functionalist approach to personhood, my focus here will be on a rather different implication, which has received far less laudatory attention. For not only does the thesis of multiple-realizability entail that persons might be realised by non-human underpinnings, it can also be taken to imply that persons, suitably tasked and arranged, might themselves realise *further persons* – i.e., that persons might find themselves to be *subpersonal* to some larger person.

In this essay I examine two instances of this Corporate View of Persons: one taken from philosophy of mind, the other from social ontology. The philosophy of mind is Daniel Dennett’s (1978) and William Lycan’s (1987) ‘homuncular functionalism’. It proposes to explain the psychological functioning of human persons by positing within them not sub-personal functions, but sub-personal *functionaries* – homunculi, or (as Lycan sometimes calls them) Lilliputians.⁹ The social ontology is Philip Pettit’s ‘theory of group agency’, which holds that suitably designed social entities – institutions, organisations, commercial corporations – should be recognised as persons in their own rights, as supra-personal Leviathans with ‘minds of their own’.¹⁰ As these epithets suggest, my own feeling is that these two views

⁹ The mention of Lilliputians appears in Lycan’s (1979) attempt to strengthen Block’s complaints against machine functionalism.

¹⁰ Though Pettit clearly distinguishes his view from Hobbes’s ‘authorisation’ theory of group agents, the broadly Hobbesian view of mind and personhood are prominent influences upon Pettit, and the famous frontispiece of *Leviathan* aptly captures the view of corporate persons I take Pettit to defend.

share a certain monstrousness, and the goal of this essay is to say just what this monstrousness consists in – without thereby inviting renewed allegations of chauvinism. The way I will put it is that the Corporate View goes wrong by coupling a genuine insight concerning personhood with a debilitating prejudice. The insight is that persons can only be understood *interpersonally* – in terms, that is, of the attitudes and activities of other persons. The prejudice is that personhood must be explained *intra-personally* – that is, in terms of persons' parts. The upshot of conjoining an interpersonal approach to personhood with an intrapersonal one is, I claim, a kind of radical alienation; though the Corporate View appears to engender a bizarre proliferation of persons, its actual effect is just the opposite – *no genuine persons* are visible from the Corporate View.

In Section 1, I delineate two 'aggregative problems of personhood' which philosophers working in social ontology and the philosophy of psychology must inevitably confront. In Section 2, I characterise homuncular functionalism and the theory of group agency as *integrative* solutions to these aggregative troubles, and this corresponds to presenting the 'genuine insight' of the Corporate View. However, this faint praise gives way, in Section 3, to my contention that the Corporate View is ultimately an alienating conception of personhood. I conclude with a tentative communitarian proposal that seeks to avoid the radical alienation without giving up on the genuine insight of the Corporate View.

1. Aggregation

Consider the following statements, the first due to a distinguished neuroscientist, and the second to a prominent sociologist:

"What you see is not what is really there; it is what your brain believes is there... Your brain makes the best interpretation it can according to its previous experience and the limited and ambiguous information provided by your eyes" (Crick 1995: 30)

"One must not form the narrow-minded idea that the petty bourgeoisie wants on principle to enforce an egoistic class interest. It believes, rather, that the special conditions of its emancipation are the general conditions through which alone modern society can be saved" (Merton 1973: 14-5)

Here we have brains and social classes described in the same intentional or ‘personal-level’ idiom – described as believing, interpreting and wanting. Such examples, which can be multiplied indefinitely¹¹, are striking not only because they extend the laurel of personhood to things from which common sense would withhold it, but because they personify things which we tend to think of as *mereologically* related to persons proper. That is, they propose to treat persons, their constituents, and the things they constitute, in roughly the same way.¹²

What should we make of such talk? It is standard to dismiss even scientists’ personifications of brains and social groups as mere *façons de parler* – harmless shorthand, perhaps, but not anything meant seriously. But the proponents of the Corporate View whose views I shall discuss here do not take this route. They contend that the intentional characterisation of sub- and supra-personal entities plays an ineliminable explanatory role in cognitive and social science (respectively). Dennett claims that “[f]ar from it being a *mistake* to attribute... intentionality to the mereological parts of persons, it is precisely the enabling move that lets us see how on earth to get whole wonderful persons out of brute mechanical parts” (Dennett 2007: 88). Likewise, Pettit insists that “talk of group agents cannot be dispensed with in favour of talk about individual agents” (List & Pettit 2011: 5). To see why such talk is purportedly both illuminating and indispensable – and hence why it should be taken seriously – we need to attend first to two familiar concerns, which I will describe as ‘aggregative problems of personhood’.

1.1 Sub-Personal Aggregation

One quite popular way of reading the story of 20th century philosophy of mind is as a persistent struggle with the difficulty alluded to by Dennett above, of getting ‘whole wonderful persons out of brute mechanical parts’. Analytic behaviourism and mind-brain identity theory, in recoiling from the Cartesian invocation of ‘para-mechanical’ (Ryle 1949)

¹¹ We are assured they are “widespread amongst contemporary neuroscientists” (Bennett & Hacker 2003), “ubiquitous in AI” (Dennett 1978) and implicit in a “great deal of explanatory work in the social sciences” (Tollefsen 2002a: 27).

¹² As a number of philosophers (e.g., Clark 1994; Tollefsen 2002b; Huebner *et al* 2010) have noted, this manner of treatment is also a feature of much of our everyday talk. For example, I can say ‘my legs are tired’ when it is I that am, or ‘the Catholic Church doesn’t believe in abortion’ perhaps referring to the claims of a spokesperson. I do not think this everyday usage contradicts my claim about the common sense reference of ‘person’, since this sort of talk is not often meant or taken literally; the point of my focussing on the prevalence of similar talk within respectable scientific discourse is to suggest that it cannot simply be dismissed as mere *façon de parler*.

parts, succeed only in *explaining away* (whole wonderful) persons, by too readily identifying them with their mechanical parts or outward performances. Later, less reductive theories of mind (like machine functionalism and anomalous monism) attempt to do justice to wonderful persons but, as a result, fail to meaningfully relate these to their brutish parts. The perennial issue is what we might call a problem of 'sub-personal aggregation': how can the parts of persons *add up to* whole persons? Though it has parts, a person is no 'mere aggregate' of its parts; there is – we seem to want to say – something special, or 'autonomous' about the whole. What is needed is a mereological account of persons that will *illuminate* but not *eliminate* this autonomy; an explanation of how a person's parts add up to the whole person, which does not thereby represent them as a mere aggregate of parts.

Homuncular functionalism is developed at a particular situs of this perennial problem – the trouble faced by functionalist theories of mind in accommodating 'phenomenal consciousness' (Chalmers 1996; Block 1995). This concern received its seminal expression in Ned Block's example of the 'homunculi-head':

"imagine a body externally like a human body, say yours, but internally quite different. The neurons from sensory organs are connected to a bank of lights in a hollow cavity in the head. A set of buttons connects to the motor-output neurons. Inside the cavity resides a group of little men... Through the efforts of the little men, the system realizes... you" (Block 1980: 276)

Block's example is intended to draw out the intuition that there is more to mind than can be functionally articulated. Specifically, the qualitative aspects (the 'what-it-is-like-ness') of mindedness are presumed to be absent from the homunculi-head, and this makes us reluctant to attribute genuine mentality to it, even though the creature is *ex hypothesi* functionally identical to a person.

Homuncular functionalism construes Block's challenge as having a peculiarly aggregative character; Lycan's paraphrase of the challenge is, "how could an actual *quale* be produced by or in a mere aggregate of dull little men shuffling index cards about" (Lycan 1987: 26). But though the challenge is focused on the production of qualia, it is concerned with them only insofar as they are *states of persons*: "the subject of... consciousness is *the person*, and not

any of the person's parts" (Dennett 1978: 150). As this seems to indicate, there is a *prima facie* tension between recognising something as a person and seeing it as a 'mere aggregate' – a tension which should, I think, give us pause to reconsider the intra-personal approach to personhood altogether. But, as we shall see, Lycan and Dennett think there is a way to circumvent this difficulty without giving up on the broadly intra-personal approach, for one can still "construct a fully-fledged "I" out of sub-personal parts" (Dennett 1978: 154) if one is prepared to personalise the sub-personal realm. Their response thus seeks to turn the tables on Block's challenge, embracing it as an exemplification rather than a *reductio* of functionalism. According to the homuncular functionalist, not only would the fictional homunculi-head, if it were real, count as a genuinely minded agent, but, in some sense, we *are all homunculi heads*.

1.2 Supra-Personal Aggregation

We will come to the details of this extraordinary-sounding claim in Section 2, but now let us attend to our second aggregative problem of personhood, encountered at the supra-personal, rather than the sub-personal, level. As observed above, it is common within social science to depict certain kinds of social groups – like classes, nations, governments, etc., – not simply as collections of individuals but as unified intentional subjects in their own rights. We might add that this way of portraying groups is also a feature of much of our everyday talk, and is enshrined in several prominent legislative frameworks, where corporate entities are recognised as *personae fictae*.

An orthodox way of interpreting such talk takes its cue from this legal concept, contending that attitude ascription to groups is always fictive, or, more precisely, elliptical:

"To ascribe mental predicates to a group is always an indirect way of ascribing such predicates to its members... With such mental states as beliefs and attitudes the ascriptions are of... a summative kind. To say that the industrial working class is determined to resist anti-trade union laws is to say that all or most industrial workers are so minded" (Quinton 1975: 17)

So perhaps, when our sociologist describes the 'petty bourgeoisie' as *believing* and *wanting*, we should understand this is as a sort of shorthand for the claim that a majority, or some

designated sub-class, of the bourgeoisie holds these attitudes. Hence, the bourgeoisie is not, strictly speaking, an entity in its own right, but is instead an aggregate or sum of its members.

This way of reinterpreting intentional ascriptions to social groups may be suitable for cases like the bourgeoisie, but recent results in the field of judgement aggregation suggest that it will not be appropriate for all kinds of social groups. Certain groups face a distinctive difficulty – which Pettit calls the ‘Discursive Dilemma’ – when they come to form their ‘group attitudes’, and this suggests that these attitudes cannot plausibly be understood in the orthodox summative fashion.

The Discursive Dilemma is Pettit’s generalisation of a problem recently noticed to affect adjudication in multi-member courts, known as the ‘Doctrinal Paradox’ (Kornhauser & Sager 1993). When a panel of judges decides a case that depends on logically interconnected issues, a seemingly plausible method of aggregating their individual judgements – namely, majority voting – can lead to unwelcome results. More specifically, even when each judge forms a set of views that is consistent with the relevant legal doctrine, a majoritarian ‘aggregation function’ may produce a set of collective views which is inconsistent.¹³

Pettit has argued that this problem is not confined to legal contexts, or to situations in which groups have to judge on interconnected issues,¹⁴ and he has formally demonstrated that it is not only majoritarian attitude aggregation which can lead to ‘paradoxical’ outcomes.¹⁵ The upshot is that, even when each individual in a social group is rationally exemplary, there is no aggregation function – no way of getting from the views of individuals on some issue to a group view on that issue – which can secure the collective rationality of the group.

One way of viewing the Discursive Dilemma, then, is as an insurmountable obstacle to the personhood (or ‘integrity’) of groups, and thus as a repudiation of the practice of ascribing attitudes to groups. If the aggregation of individual judgements leads inevitably to failures of rationality at the group level then we might conclude that “the possibility of a paradoxical

¹³ The Doctrinal Paradox is discussed at length in Essay 2, and for that reason the standard example used to illustrate the paradox is not reproduced here (see pp. 36-7 for this example).

¹⁴ Essay 2 seeks to cast doubt upon these two generalising moves.

¹⁵ This is demonstrated in List and Pettit’s (2001) impossibility theorem on the aggregation of judgement.

distribution of views amongst the members of the group confounds the pursuit of [group] integrity” (Kornhauser & Sager 2004: 252). But Pettit’s insistence that groups in paradoxical scenarios face a certain kind of *choice* makes another option available – one which seeks to turn the tables on the debunkers of group personhood. Pettit’s route out of the dilemma is to agree that group attitudes cannot be aggregatively related to the attitudes of group members, but to say that this *vindicates*, rather than undermines, the practice of ascribing attitudes to groups. This is because groups are not *bound* to be fully responsive to individuals: they can opt to make themselves collectively rational, and in so doing they may develop ‘minds of their own’. Crucially for Pettit, the failure of group attitudes to be readily reducible to individual attitudes makes it *all the more appropriate* to conceive of groups as persons, since it secures a certain autonomy for group persons from their underpinnings.

2. Integration

In the previous section we encountered two ‘aggregative puzzles of personhood’: one a perennial problem for philosophers of mind, the other an increasingly central concern within social ontology. The first is the problem – promulgated originally as a *reductio* of functionalism – of how a person’s conscious experience could possibly be the product of ‘a mere aggregate of little men shuffling index cards about’ (since the functionalist is seemingly committed to saying that it could). The second is the problem of attitude aggregation exemplified by the ‘Discursive Dilemma’, a difficulty which might be thought to compromise our practice of personifying groups by ascribing attitudes to them. However, as we have also seen, each of these puzzles has been recognised by the proponents of the Corporate View “not so much as a problem... but as an opportunity” (Kornhauser & Sager 2004: 254). The homuncular functionalist embraces Block’s example, saying that the conscious experience characteristic of persons not only *might be* but *actually is* the product of homuncular activity. Similarly, the theorist of group agency welcomes the Discursive Dilemma, discerning in the failure of straightforward attitude aggregation the promise of a fuller, ‘non-redundant’ (because non-summative) realism about group persons.

In this section I examine some of details of these two table-turning proposals, showing that both can be seen as *integrative* responses to the aggregative challenges.

2.1 Sub-personal Integration

Drawing on research methodologies endorsed in cognitive psychology¹⁶ the homuncular functionalist proposes that:

“we view a person as a corporate entity that corporately performs many immensely complex functions – functions of the sort usually called mental or psychological... On this view the psychological capacities of a person and the various administrative units of a corporate organisation stand in functional hierarchies of just the same type and in just the same sense” (Lycan 1987: 40)

The idea is that we should understand a person in the same way that we understand an institution: we see her as internally constituted by a collection of sub-agencies distributed across a hierarchy of organisational levels. Agencies on these organisational levels are staffed with *Lilliputians*: tiny bureaucrats postulated ‘formally’ (Dennett 1978: 123) as ‘theoretical entities’ (Lycan 1987: 40). The top brass in the organisational structure serve the purposes of the holding person herself, heading up the immediately subpersonal-level agencies and departments¹⁷ that enable her to get on in the world – to *personate*. Each of these agencies has, in turn, its own sub-organisational make-up that sub-serves its purpose, and this same pattern continues on down the hierarchy, until at some point the labour required is so menial that ‘sub-persons’ can be replaced by brute mechanisms. At this point the institution bottoms out into neurophysiology: the theoretically posited Lilliputians, whose complex collaborative efforts realise the mental life of the holding person, are revealed to be nothing but mechanistic states and events.

This ‘top-down’ (Lycan 1987: 38; Dennett 1978: 110) approach has a number of distinctive advantages. For one thing, it allows homuncular functionalism to avoid the charge of explanatory emptiness traditionally associated with the invocation of homunculi.¹⁸ As Lycan explains:

¹⁶ By, for example, Attneave (1960), Fodor (1968), and Minsky (1985)

¹⁷ E.g., a PR department, an information gathering agency, a data centre, an executive branch, and so on (Lycan 1987: 40-41; Dennett 1978: 123).

¹⁸ The seminal expression of this charge is due to Skinner: “Unable to explain how or why the person we see behaves as he does, we attribute his behaviour to a person we cannot see...[but] about whom we are not inclined to ask questions... The function of the inner man is to provide an explanation which will not be explained in turn” (Skinner 1971: 14)

“we explain the successful activity of one homunculus, not by idly positing a second homunculus within it that successfully performs that activity, but by positing a team consisting of several smaller, individually less talented and more specialised homunculi – and detailing the ways in which the team members cooperate in order to produce their joint or corporate output” (Lycan 1987: 40)

Moreover, the invocation of a Lilliputian taskforce as intermediate between persons and their mechanical parts enables homuncular functionalism to avoid the *explaining away* tendency exhibited by crude reductionist theories of mind. The homuncular functionalist diagnoses this tendency as the product of a mistaken inference – the inference from the commitment to explain persons *sub-personally* to the requirement that persons be explained immediately and exclusively in *impersonal* terms (in terms of brain states, perhaps, or behaviour). By populating the sub-personal realm with intentional systems, Dennett and Lycan can accommodate both the well-worn negative point that “mental events... resist capture in the nomological net of physical theory” (Davidson 1980: 137), and the increasingly fashionable positive contention that personhood is essentially a ‘relational’ or social phenomenon, all within a scheme that, in the end, sees persons as nothing beyond their neurophysiological underpinnings.

Most saliently for my purposes here, the top-down approach suggests a way of overcoming the puzzle of subpersonal aggregation described above. From the perspective of homuncular functionalism, the Lilliputian bureaucrats who realise the psychological functioning of a person are not just constitutive *parts*, they are co-operating *participants*, postulated just to collectively subserve some higher-level purpose. Taking the top-down approach enables us to view lower-level constituents in this distinctively teleological way, and this militates against the thought driving the aggregative puzzle – the thought, roughly, that all mereological explanation impugns. When a person is seen to be realised by a range of nested subsystems collaborating for the purpose of realising her, the temptation to view her as a *heap* of these systems (or, since the systems bottom out into neurophysiology, as a *bundle* of neurons) is radically undermined. What the distinctive, top-down style of explanation characteristic of

homuncular functionalism provides is a way of talking about persons such that they are unified, incorporated entities: no longer ‘mere aggregates’, but, instead, ‘integrates’.¹⁹

In this way homuncular functionalism holds the promise of a kind of integrative response to the aggregative puzzle with which we began. But we may still wonder whether homuncular functionalists are fully *entitled* to postulate inner agents in this way. After all, nobody is claiming to have *discovered* these tiny bureaucrats, and positing them is certainly not explanatorily economical or elegant. We might wonder whether the homuncular strategy is of a kind with what Russell famously derided as “the method of ‘postulating’ what we want” – a method with all of the “advantages of theft over honest toil” (Russell 1919: 71). The postulation of homunculi may not be idle but is it *honest*?

The question of what entitles the homuncular functionalist to postulate these inner bureaucrats is of the utmost importance for the claims I wish to make in this essay. The answer comes in the form of Dennett’s interpretationist approach to intentionality, according to which there is nothing more to *actually* having a belief, or to being a *genuinely* intentional creature or system (a ‘true believer’), than being reliably and profitably interpretable as such, from the ‘intentional stance’.²⁰ When interpreting from the intentional stance, an interpreter assumes that their subject has a particular kind of psychological repertoire (including, paradigmatically, beliefs and desires), that she is basically rational (she forms beliefs favoured by evidence, or entailed by her other beliefs, and forms intentions to act in line with her beliefs and desires), and that she is not generally akratic (she generally does act in line with her intentions). To the extent that these assumptions *pay off* by making the subject ‘usefully and voluminously predictable’ (Dennett 2009: 339), the subject is a genuine intentional system.

The Lilliputians postulated by homuncular functionalism are nothing more or less than intentional systems in this sense. But the fact that interpretation from the intentional stance

¹⁹ It is important to note that all of this is only the groundwork for a homuncular account of consciousness, which both Dennett (1978: Ch. 9) and Lycan (1987; 1996) undertake to provide. Nevertheless, if Block’s example is meant to challenge the very possibility of such an account, then, on one level, this challenge has been answered.

²⁰ See esp. Dennett (1978, Ch. 1; 1987, Ch. 2)

is tied to useful and voluminous prediction ensures that the postulation of subpersonal intentional systems engendered by this approach is neither 'idle' nor an act of 'theft'. Instead it is the kind of honest toil involved in any process of reverse engineering: subsystems of a certain sort are theoretically posited, but these posits are not made by fiat so much as by hypothesis. Assumptions about their character are constantly being tested and adjusted in light of what they lead us to predict; Lilliputians might be originally be posited on the basis of 'what we want' but they cannot on that same basis be retained.

2.2 Supra-Personal Integration

Let us turn now to the supra-personal case of social groups. The problem of aggregation encountered at this level related to group attitudes, and was exemplified by the 'Doctrinal Dilemma'. The problem, in short, was that deriving group-level attitudes in a summative fashion (by, for instance, majority-voting on all issues) could result in a set of collective attitudes which would not be rational to hold. For certain groups, like the panel of judges in the original judicial example, this seemed like a problem, and Pettit's route out of this problem was to say that groups like this should strive to collectivise reason. The upshot of collectivising reason is that the group achieves a kind of integration:

"Instead of speaking of groups that collectivise reason... I shall talk... of integrations of people, of integrated collectivities, and of social integrates. This way of speaking sounds a contrast with those groups that... do not impose the discipline of reason at the collective level. These we naturally describe as aggregations of people, as aggregated collectivities, or just as aggregates. [I] argue that... social integrates are going to be intentional and personal subjects" (Pettit 2002: 454)

We might wonder just how groups should set about collectivising reason in order to become 'integrates'. Pettit thinks that there are several procedural mechanisms a group might adopt in order to ensure that their group-level attitudes will meet the demands of rationality, so let us examine one of these to see how it effects a certain kind of *integration*.

Consider the simple 'premise-based' approach, which Pettit favours as a solution to the standard rendition of the Doctrinal Paradox. The premise-based approach requires judges to identify certain issues as logically dependent on certain others: in the standard example

(described on pp. 36-7), the issue of the defendant's liability is seen as depending on two 'prior' issues: whether a contract existed, and whether the defendant performed some action. The benefit of the premise-based approach is that judges need only adjudicate issues which have logical priority, allowing *modus ponens* to do the rest (thus ensuring group-level consistency).

Now of course the simple premise-based approach has certain deficiencies – just which issues are logically prior will not always be obvious²¹, the relations between issues will not always suitably strict²², and the procedure is perhaps too rigid to allow proper sensitivity to the all-round demands of reason²³ – but these needn't preoccupy us now. Instead let us consider the implications of adopting such a procedure for the relationship between group-level attitudes and the attitudes of group members. Pettit cashes out these implications by contrasting two ways in which a group's attitudes might *supervene* on the attitudes of group members. He observes that the collective rationality secured by the premise-based approach is in conflict with 'proposition-wise supervenience' (List & Pettit 2011: 67-8) – a supervenience relation between the group-attitude on some issue and the attitudes of group members on that issue. This is illustrated by Profiles 1 and 2 below, in which operating under a premise-based procedure reaches a different outcome (i.e., a different decision on the liability of the defendant) even though the judges' *individual* judgements on the outcome are identical.

But even though employing a procedure such as the premise-based approach to collectivise reason puts paid to proposition-wise supervenience, it is still the case that a 'holistic' relation of supervenience obtains between group attitudes and the attitudes of individuals (List & Pettit 2011: 69). In other words, the *whole set* of group attitudes will still depend on the underlying *whole sets* of attitudes belonging to individual members. The profiles below illustrate this too: the difference in group-level attitudes that can be discerned between them can be explained in terms of the difference across the two profiles in Judge B's judgment on Issue 2 (the question of whether there was a breach). The 'change' in B's attitude on this issue results in a concomitant alteration in the group's attitude on this issue, and hence (since a

²¹ As List and Pettit (2011: 57) put it, "One person's premise is another person's conclusion".

²² Brandom (2012: 2) claims that "Legal reasoning... is seldom formally, logically valid... almost all the reasons considered are defeasible".

²³ This is discussed in Pettit (2007).

premise-based approach is being deployed) on the outcome of the case. Set-wise supervenience ensures that this must always be the case: no change can occur in the group-level attitudes without some change, somewhere, in the underlying base of individuals' attitudes (conceived as a whole).

Profile 1

Judge	Contract?	Breach?	Liable
A	Y	N	N
B	N	Y	N
C	Y	Y	Y
Premise-based Approach	Y	Y	→ Y

Profile 2

Judge	Contract?	Breach?	Liable
A	Y	N	N
B	N	N	N
C	Y	Y	Y
Premise-based Approach	Y	N	→ N

This distinction between proposition-wise and set-wise supervenience provides a useful way of capturing Pettit's response to the aggregative challenge posed by the Discursive Dilemma. The challenge was to show how individual attitudes could add up to group attitudes without impugning the autonomy of the group attitudes thus derived, since we want to avoid the conclusion that group attitudes are a 'mere aggregate' of individual attitudes. Now we can see that by deploying an organisational mechanism designed to collectivise reason – for instance, the premise-based approach – the desired solution can be provided. Group attitudes depend on nothing besides the attitudes of group members but the failure of proposition-wise attitude aggregation makes it difficult to label these group attitudes 'mere aggregates' of the attitudes of individuals.

In this way the inculcation of a decision-making mechanism such as the premise-based approach enables the group to become a ‘social integrate’ rather than an ‘aggregated collectivity’. But what entitles Pettit to say further that, ‘social integrates are going to be intentional and personal subjects’? The answer, it turns out, is similar to that of the homuncular functionalist. We are entitled to recognise certain groups as rational agents or persons because seeing them in this way opens up a new level of understanding and prediction – i.e., some groups are appropriate subjects of the *intentional stance*:

“The shift in perspective opened up by the recognition of group agents is parallel to the one opened up... when we move from thinking of an organism as a purely biological system to viewing it as an intentional agent... In short, it parallels the move from taking a ‘physical stance’ towards a given system to taking an ‘intentional stance’... Although their agency depends on the organisation and behaviour of individuals, as individualism requires, [group agents] display patterns of collective behaviour that will be lost on us if we keep our gaze fixed at the individual level” (List & Pettit 2011: 6)

3. Alienation

The two views sketched above – homuncular functionalism and the theory of group agency – are both recognisably versions of what I earlier called the Corporate View of Persons. That is because both are content to view certain persons (individual human beings in the case of homuncular functionalism; group agents in the case of the theory of group agency) as *constituted* by other, smaller persons. We have also seen that both views strive to secure a kind of autonomy for these corporate persons, and in each case this is done by venturing an integrative solution to an aggregative challenge. The move from aggregation to integration seems to allow the proponents of the Corporate View to accommodate an important insight regarding personhood – that persons are only intelligible in terms of the attitudes and activity of other persons – within a broadly reductionist, or ‘intra-personal’ approach.

In this final section I want to pursue a particular line of objection against homuncular functionalism and the theory of group agency, one which articulates what I take to be the dark side of their shared strategy. My focus will be on the ‘intentional stance’ approach to

intentionality and personhood which both theories rely upon to license their postulation of sub- and supra-persons, but which I think does violence to our concept of personhood.

3.1 Nonsense

One objection to the liberal use of intentional idiom within neuroscience and cognitive psychology draws inspiration from a claim made by Wittgenstein:

“Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious”
(*Philosophical Investigations*, S281)

The most forceful contemporary exponents of this complaint are Bennett and Hacker (2003), who contend that neuroscientists and cognitive psychologists who lapse into talk of brains *knowing, thinking, inferring* etc, are guilty of a ‘mereological fallacy’. That is, they mistakenly attribute to a part of a person some capacity or property which can only rightly be attributed to the whole person – and the result, according to Bennett and Hacker, is nonsense:

“Our point... is a conceptual one. It makes no sense to ascribe psychological predicates to the brain... The resultant combination of words does not say something that is false, rather it says nothing at all, for it lacks sense... psychological predicates are ascribable [only] to the whole animal, not to its constituent parts” (Bennett & Hacker 2003: 72)

Of course, if we accept that only a ‘human being’ (Wittgenstein) or ‘whole animal’ (Bennett & Hacker) is the proper subject of intentional or personal-level characterisation then we will need to admit that the ascription of beliefs and actions (and so on) to social groups is another instance of this conceptual confusion. Once again the charge can be construed as one of fallacy – this time not the mereological fallacy, but the *fallacy of composition*. In other words, to talk of the group as a believer or an agent is to erroneously transfer to the whole group some properties belonging properly only to parts of the group – i.e., its members.

Do such allegations of nonsense threaten the very core of the Corporate View? It is hard to see how they could without the complainants saying a lot more about just why the intentional idiom *cannot* be extended towards sub- and supra-personal entities. For all the fallacies show us is that it is not *generally valid* to transpose properties from whole to parts or

from parts to whole; they do not in any way suggest that parts and whole could not share properties. But homuncular functionalists and their neuro- and cognitive-scientific counterparts, just like theorists of group agency and their social scientific counterparts, may have powerful independent reasons for positing their Lilliputians and Leviathans.

Our brief exploration of homuncular functionalism and the theory of group agency provided a glimpse of just what these reasons might be. As Dennett points out, it is possible to take the Wittgensteinian point in precisely the opposite way to Bennett and Hacker – i.e., as *support* for the more liberal extension of intentional idiom:

“I take myself to be *extending* Wittgenstein’s position: I see that... brains and their parts *do* ‘resemble a living human being (by behaving like a human being)’ – and this resemblance is sufficient to warrant an adjusted use of psychological vocabulary to characterise their behaviour” (Dennett 2007: 78)

This seems enough to show that accusations of *fallacy* and *nonsense* against the Corporate View are vastly overstated. In the absence of an argument expressly precluding the extension of psychological vocabulary to sub- and supra-personal entities, these accusations seem to be premised on the thought that because this extension is ‘liberal’ it must be unprincipled – and this is surely a sort of conservative prejudice or ‘chauvinism’.

3.2 Violence

We may still, however, wish to know in what ways parts of persons, or, for that matter, conglomerations of persons, ‘resemble’ living human beings, and in what sense the psychological vocabulary used in their characterisation is ‘adjusted’. On this Dennett has more to say:

“We don’t attribute *fully fledged* belief... to the brain parts... we attribute an attenuated sort of belief stripped of many of [its] everyday connotations... Just as a young child can *sort of* believe that her daddy is a doctor... so... some part of a person’s brain can *sort of* believe that there is an open door a few feet ahead” (Dennett 2007: 87)

As we saw above, Dennett's intentional stance approach to intentionality (relied upon by both homuncular functionalism and the theory of group agency) sees the fact of whether something is an intentional subject as dependent on whether it pays to view them as such. The fact that interpretive payoff is variable (different sorts of subjects will be *more or less* predictable from the intentional stance) makes room for talk of fully-fledged and less-than-fully-fledged (or 'sort of') belief. More generally, it makes the extent to which subjects can appropriately be thought of as intentional systems a matter only of degree:

"One may be tempted to ask: Are the subpersonal components *real* intentional systems? At what point in the diminution of prowess as we descend to simple neurons does real intentionality disappear? Don't ask. The reasons for regarding an individual neuron... as an intentional system are unimpressive, but not zero, and the security of our attributions at the highest levels does not depend on our identifying a lowest level of real intentionality" (Dennett 1998: 362)

Dennett's brusque 'Don't ask' reflects a confidence that we can happily eschew the 'tempting' line of inquiry which culminates in the question of whether an individual neuron is an intentional system, since either way – whether we conclude that it is, or that it isn't – our understanding of *ourselves* as intentional systems will not be thereby undermined. This is because (presumably) the reasons we have for viewing creatures like ourselves from the intentional stance are quite independent of the case of the individual neuron, and are not at all 'unimpressive'. But Dennett's confidence betrays a misapprehension of the proper point of pursuing that tempting line of inquiry. For the point of asking after the possible intentionality of individual neurons is not necessarily to insinuate that their intentionality, were it to be granted, would somehow *cheapen* – make less 'secure' – *our* intentionality. Instead, the purpose could be to insist that the view of intentionality which sees it as the product of an interpretive stance does away with any notion of *genuine* intentionality. On Dennett's view all intentionality is derived intentionality. It is only ever *as if* something was an intentional system, and organisms – as well as their constituents and the things they constitute – can only ever have *sort of* intentional states:

'By treating subsystems *as if* they had intentional properties, we are to explain why it should be that systems comprised from them should be as if *they* had such properties... it is only ever *as if* persons had such properties' (Hornsby 2000: 9)

It is this instrumental approach to intentionality which I think threatens to do violence to our concept of personhood. To see how, it is worth noting that Dennett's development of the 'intentional stance' has been seen by some philosophers²⁴ as eliding an important distinction he popularised in earlier work, between the subpersonal- and personal-levels of explanation. In *Content and Consciousness* Dennett illustrates this distinction with the example of a person's feeling pain:

"When we have said that a person has a sensation of pain, locates it and is prompted to react in a certain way, we have said all there is to say within the scope of this vocabulary... If we look for alternative modes of explanation, we must abandon the explanatory level of people and their sensations and activities and turn the sub-personal level of brains and events in the nervous system" (Dennett 1969: 93)

The point of insisting on a personal / subpersonal distinction is to register the fact that there are two quite different strategies available for understanding the behaviour of human beings. One strategy represents such a creature as part of the causal order of nature as it is represented by the physical sciences, and hence focuses on features of their subpersonal make-up which make their causal transactions with their physical environment intelligible in terms of natural laws. The other strategy represents these creatures as 'rational agents' or 'persons', as subjects whose distinctive psychological attitudes have distinctively *normative* implications:

"the concepts of the propositional attitudes have their proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be. This is to be contrasted with a style of explanation in which one makes things intelligible by representing their coming into being as a particular instance of how things generally tend to happen" (McDowell 1986: 389)

Drawing a sharp personal / subpersonal distinction is thus to insist on something which was part of the *motivation* for the homuncular approach to mind, the point that "there is no

²⁴ Notably Hornsby (2000)

prospect of finding a person intelligible in terms of physical goings on inside her” (Hornsby 2000: 3). This was seen as the principal defect of crude reductionism, and the homuncular functionalist’s idea for avoiding this defect was to find a way of characterising the inner goings-on of a person in terms that are not (or are not immediately) physical. Thus homunculi get posited as theoretical entities underpinning the top-level capacities of a person, and in this way the person is made intelligible.

But notice that, if we are committed to keeping the personal / subpersonal distinction intact, then the homuncular functionalist’s strategy will not actually pass muster. For in recoiling from the crude reductionist’s grievous error of recruiting subpersonal states and events to do personal-level explanatory work, the homuncular functionalist has been led to commit an opposite, though no less serious, transgression. The homuncular functionalist appears to be attempting to get personal-level concepts to do work in subpersonal spaces, and if this is to be allowed then the sharp distinction between personal and subpersonal levels of explanation has certainly been forsaken.

The transition in Dennett’s thinking marked by his move from insisting on the personal/subpersonal-level distinction to his liberal use of the intentional stance corresponds to a dramatic shift in his conception of a person.²⁵ For now there is no longer any ‘special’ sort of explanation reserved for persons that makes them intelligible in a distinctive, normative way, as striving to be ‘as they rationally ought to be’. Though the intentional stance involves ‘rationality assumptions’ the benchmark for success in the explanations adduced from this stance is not the distinctively normative sort of intelligibility but is instead ‘useful and voluminous prediction’. It is only ever *predictively useful* to treat certain things as though they were striving for rationality; there is no possibility that some creatures are *inherently* more suited to this sort of characterisation than others by being genuinely intentional, rational animals. By making rationalising explanation subservient to predictive explanation (the form of explanation in which things are treated as particular instances of how things generally tend

²⁵ In fact Dennett is somewhat equivocal on the conceptual relations between personhood and intentionality. In some places intentionality and personhood are used almost interchangeably (as when he refers to theoretically-positing homunculi as ‘sub-persons’), but elsewhere (notably ‘Conditions of Personhood’, in Dennett 1978) he is careful to distinguish intentionality from fully-fledged personhood, saying that being predictable from the intentional stance is a necessary but not sufficient metaphysical condition of personhood.

to happen), the intentional stance threatens to relegate the only sort of explanation from which genuine persons are ever visible.

Thus, by invoking the intentional stance, the proponents of the Corporate View discussed here commit to viewing persons of human size and shape as persons in just the same ‘derived’ sense as they view their theoretically-posited counterparts, the Lilliputians and the Leviathans. Though human persons may be, in some sense, more ‘fully fledged’ than subpersonal cognitive processes or supra-personal corporations, there is no sense in which they are more genuine.

Conclusion: A Communitarian Approach to Persons

The previous section attempts to justify the claim that, insofar as it appeals to the view of intentionality associated with Dennett’s ‘intentional stance’, the Corporate View of Persons has no place for the notion of *genuine* persons. This is why I have described the Corporate View as ‘monstrous’ and ‘alienating’: it involves a fundamental recalibration of the way we conceive of ourselves. Such revisionism is not unthinkable, but its general uptake would, I think, correspond to a kind of loss or harm – so it is better construed as *violence* than as *nonsense*. But this way of construing the problem leaves me with the burden of saying why the alienating conception of personhood is not in fact compulsory, because although nonsense might never be required, violence sometimes is.

Though I cannot hope to fulfil this burden here, I would like to end off with a brief sketch of an alternative that is not obviously deficient, and so may diminish the sense that the alienating view is compulsory. The alternative takes issue with something the proponents of the Corporate View treat as an inescapable demand on any account of persons – what we might call the demand for lower-level explanation. In the philosophy of mind this usually takes the form of a commitment to physicalist explanation, to providing an account that successfully (though perhaps gradually) connects ‘whole wonderful persons’ to their ‘brute mechanical parts’. In social ontology the core commitment is typically to some form of individualism: group persons must be shown to be the product only of individual attitudes and actions and nothing more, even if they are not ‘readily reducible’ to them. Behind these

commitments lies the conviction (or, as I see it, the prejudice) that without a lower-level explanation no *genuine* account of persons has been put forward.

What I have been arguing, then, is that if this is what we are after then genuine persons will never be genuinely explained. For it seems to be part of the concept of personhood that persons are somewhat autonomous from their underpinnings, and this introduces a tension that not even the ingenious personalising strategy of the Corporate View can eradicate. In the end, accepting the demand for lower-level explanation means giving up on the notion of genuine persons, and this seems to involve a radical and regrettable revision of how we conceive of ourselves.

How can this gloomy conclusion be avoided? One option is to eschew the attempt to explain personhood at all: we might wish to say that persons – being presupposed by all our thought and talk – are somehow ‘basic’, and so do not stand in need of explanation. Another option is to eschew only the demand for lower-level explanation; but then, if this second option is not to collapse into the first, what should we put in its place? A perfectly promising answer has been, as it were, staring us in the face: we should explain persons in terms of other persons, not by positing sub- and supra-persons, but by seeing the concept of personhood as inextricably tied to the concept of community. We should replace the vertical account with something horizontal.

How would this horizontal sort of explanation work? One plausible way of filling in the details sees a person’s personhood as constituted in part by the attitudes of their fellow persons; personhood is then not a matter of being *realised* in a certain way but is instead a matter of being *recognised* in a certain way:

“What institutes normative statuses is *reciprocal recognition*. Someone becomes responsible only when others *hold* him responsible, and exercises authority only when others *acknowledge* that authority. One has the authority to *petition* others for recognition, in an attempt to become responsible or authoritative. To do that, one must recognize others as able to *hold* one responsible or *acknowledge* one’s authority. This is according those others a certain kind of authority. To achieve such statuses, one must be recognized by them in turn” (Brandom 2009: 70)

Might this model of reciprocal recognition give us an account of personhood which is interpersonal but not intra-personal? Perhaps, though a lot more would have to be said. Would its adoption make the status of personhood somehow 'empty' – a “free-floating honorific” (Dennett 1978: 268) devoid of substantive content? Certainly not, for as Brandom (1994; 2002; 2009) has forcefully demonstrated, the invocation of community allows us to make sense of the normative statuses in terms of a range of communally-accepted and communally-enforced obligations and entitlements. Does the fact that the communitarian approach sees personhood as attitude-dependent make it susceptible to the same problem incurred by Dennett's intentional stance, whereby all intentionality becomes *as-if* intentionality? Seemingly not, since recognition, unlike interpretation, is inter-subjective, and so personal status will depend on a *pattern* of attitudes distributed across a range of individuals. In this way personhood is represented not as an intrinsic property of some creatures, nor as a robust disposition to be predictable under a range of assumptions, but as distinctive kind of social achievement, carrying with it a range of distinctively social implications.

Finally, if the community of persons is delimited by attitudes of recognition, does this not then inevitably lapse into another kind of exclusionism – not chauvinism, perhaps, but a kind of cliquism? I do not think so, since on the communitarian view the community of persons is never decisively closed off to anyone who could in principle participate in it: the only characters automatically excluded are those who cannot even strive to be recognised as community members. The *possible* personhood of sub- and supra-personal entities would thus depend on whether they can strive for personhood, and their *actual* personhood would depend on whether their striving is rewarded by reciprocal recognition. I remain agnostic on both of these issues, but as a policy I think this is neither excessively liberal nor unfairly chauvinist; it seems to me to gets things just about right.

Essay 2: Doctrine and Reason²⁶

Introduction

The Doctrinal Paradox – first identified in jurisprudence as a ‘quirk’ of multi-judge tribunals – has over the past decade received a considerable amount of attention from philosophers interested in judgement aggregation.²⁷ This attention is due in no small part to the efforts of Philip Pettit, who argues that the paradox is far more general, and consequently has far greater significance, than the jurisprudential literature suggests. Pettit claims that the difficulty faced by collegial courts is only an instance of a broadly ‘Discursive Dilemma’, “[which] is not tied to the acceptance of common doctrine, only to the enterprise of making group judgements on the basis of reasons” (Pettit 2001: 272).

According to Pettit, the dilemma in this general form has important consequences for social ontology and democratic theory.²⁸ These claims have been met with some resistance,²⁹ but what to my mind is most curious about the ensuing discussion has been the almost exclusive focus on how the paradox should be resolved, and (relatedly) what its significance is, while the generalising move upon which such debate rests has commanded comparatively little attention. In light of this, I wish to explore and raise doubts about the extent to which the scope of original jurisprudential difficulty can be broadened in the manner envisaged by Pettit. In short, I hope to defend the Doctrinal Paradox over the Discursive Dilemma.

Since this is my aim, it is worth being clear from the outset on what is meant by ‘generality’ or ‘scope’ here. At first blush, this appears simple: the level of generality of the Doctrinal Paradox refers to the range of contexts in which the paradox is applicable. But ‘applicability’ admits of two different interpretations – one descriptive, the other normative. On the descriptive understanding, fixing the level of generality is a matter of isolating the structural features of paradoxical cases and then seeing in what kinds of contexts that structure might

²⁶ I wish to thank Prof. Robert Brandom for granting me permission to cite his unpublished paper, ‘A Hegelian Model of Legal Concept Determination’

²⁷ Recent discussions from the perspective of legal philosophy include Chapman (2002), Kornhauser & Sager (2004), Nash (2003), List & Pettit (2005); from the perspective of social epistemology, see Pettit (2003), Goldman (2004), List (2005); for a good survey of the many formal social-choice-theoretic contributions, see List & Puppe (2007).

²⁸ See esp. Pettit (2001, 2004), List & Pettit (2011)

²⁹ Notably from Kornhauser & Sager (2004)

be instantiated. Descriptively, the paradox has broad scope just if it can crop up in a broad array of contexts. On the normative understanding, what is at issue is the *appropriateness* of the conditions required for the paradox to be instantiated: whether the sort of reasoning which leads to paradox is something to which everyday reasoners should aspire. This can be separated from the question of whether and where those conditions obtain.

In light of this distinction, I can register my first complaint against Pettit: he fails to disambiguate these two senses of applicability, the upshot being that, wherever he can conceive of the paradox being instantiated, he appears to hold that it is appropriate that it be so. My suspicion is that this is no oversight but a symptom of a tendency to hold up as an ideal a certain doctrinal conception of the exercise of reason. I can also restate my thesis: it might be that the Doctrinal Paradox can be replicated in a variety of contexts (I remain agnostic on this issue), but, where it can, apart from a few special contexts (legal adjudication may be one), the presence of the conditions which give rise to the paradox amount to a subjugation and not an exemplification of reason.

In what follows I lend partial support to this thesis by arguing that Pettit's best efforts to generalise the Doctrinal Paradox do not succeed in showing how it can survive without the presence of some kind of exogenous constraint or 'doctrine' to which the individual reasoners comprising a group are beholden. This means that, for all Pettit has argued, the paradox and its consequences will not be broadly discursive unless we conceive of reason itself in doctrinal terms. Though I think there are good grounds for rejecting a doctrinal notion of reason, I will not explore this in much detail here; I take it that Pettit seeks to avoid (or at least not rely upon) such a conception, and in this I think he fails. In Section 1, I rehearse the original, jurisprudential version of the paradox; in Section 2, I trace three explicit generalising moves made by Pettit (2003) and argue that not one is compelling. Finally, in Section 3, I consider whether the doctrinal conception of legal reasoning assumed by both Pettit and his opponents is appropriate, and gesture towards a non-doctrinal alternative.

1. The Doctrinal Paradox

The Doctrinal Paradox was first introduced into jurisprudential literature by Kornhauser and Sager (1986, 1993), who noticed that, "when multi-judge tribunals confront cases with

multiple issues, the protocol for deciding cases can become quite problematic” (Kornhauser & Sager 1993: 11). ‘Cases’ here refers to the verdicts or outcomes decided by judges (‘guilty’ or ‘not guilty’, for example), while the ‘issues’ are the salient factors which judges must consider in arriving at their verdicts. What enables judges to identify which factors are salient, as well as how to relate those factors to cases, is *legal doctrine* – the resources within the legal system which guide the decisions of judges.

Kornhauser and Sager illustrated the paradox with a simplified example in which a defendant is being sued for damages arising from an alleged breach of a contract. Legal doctrine dictates that the decision of the court – the decision as to whether or not the defendant is liable – must depend on just two issues: whether a contract existed, and whether the behaviour of the defendant was such that it would constitute a breach of contract (a material breach). More precisely, if and only if a contract existed and there was material breach, should the judges decide ‘liable’; otherwise they should decide ‘not liable’. Such a case becomes ‘paradoxical’ if the views of the judges are as follows (Kornhauser & Sager 1993: 11):

Judge	Issue 1: Contract?	Issue 2: Breach?	Outcome: Liability?
A	Y	N	N
B	N	Y	N
C	Y	Y	Y
Majority	Y	Y	N

The reason this profile is seen as ‘paradoxical’ is that the bottom line – the view of the majority – is inconsistent with legal doctrine, and so the decision-making protocol adopted by the court will make a crucial difference to the overall verdict. If the court allows a majority of outcome views (i.e., verdicts on the matter of liability) to determine the overall verdict – what Kornhauser and Sager calls the ‘case-by-case protocol’ – the defendant will be found to be not liable. If instead they apply legal doctrine to the majority-favoured issue views (i.e., verdicts on the contract and breach issues) – the ‘issue-by-issue protocol’ – then the defendant will be judged to be liable.

According to Kornhauser and Sager, the situation is stubbornly paradoxical because not only does the legal system³⁰ happen not to favour one protocol over another, or provide a rule dictating when each protocol is to be adopted (deficiencies that could be corrected), but because any such stipulation would be bound to get things wrong in certain cases. They maintain that certain cases ought to be resolved via case-by-case voting, while others ought to be resolved by tallying judges' views on issues, and, importantly, there is no antecedent way to distinguish the one kind of case from the other. Their preferred mechanism for escaping the paradox is thus a 'metavote': upon encountering a paradoxical case, the court votes on which protocol to adopt in the circumstances. This allows judges to display sensitivity to features of the particular case that favour one voting protocol over another.

2. The Discursive Dilemma

It should be clear that paradoxical cases will arise for multi-judge tribunals even when things differ slightly from the original example. There could be more judges, and more issues. The paradox requires only that a majority assent to each of the issues but that the sub-group comprising those who assent to both issues is a minority. Also, the way that the issues connect with one another to determine the outcome may differ: the relevant legal doctrine might connect issues with outcome in some way other than simple conjunction – disjunction for instance, or some more complex logical relation. But these are just variations of Kornhauser and Sager's Doctrinal Paradox; they do not at this point support Pettit's contention that the trouble facing the court is an instance of something more general.

In 'Groups with Mind of their Own' (2003) Pettit presents his case for talking of a Discursive Dilemma rather than a Doctrinal Paradox, contending that the difficulty facing collegial courts generalises along both contextual and logical lines. Paradoxical cases are not restricted to legal or judicial contexts – instead, they will arise for "any groups that seek to make deliberative, reasoned judgements" (Pettit 2003: 175). Moreover, the court's difficulty is not just a choice between two voting protocols, but something grander – the choice of "whether to let the views of the collectivity on any issue be fully responsive to the individual views of members... or whether to ensure that the views of the group are collectively rational" (Pettit

³⁰ The system to which Kornhauser and Sager (1986, 1993) refer directly is appellate adjudication in the USA.

2003: 174). In support of broader applicability Pettit presents two sets of considerations – the ‘Social Generalisation’ and the ‘Diachronic Generalisation’ – while in support of a broader characterisation of the nature of the problem, he presents the ‘*Modus Tollens* Generalisation’. Let us consider each of these in turn.

2.1 The Social Generalisation

The social generalisation aims to show that the structure of the Doctrinal Paradox is not restricted to adjudicatory settings:

“A paradox of the sort illustrated will arise not just when legal doctrine dictates that certain considerations are conceptually or epistemically prior to a certain issue... It will arise whenever a group of people discourse together with a view to forming an opinion on a certain matter that rationally connects, by the lights of all concerned, with other issues... There are many social groups that have to make judgements on various issues and that routinely do so by reference to considerations that are privileged within the group” (Pettit 2003: 170-2)

On the face of it, this seems plausible both as a description of the operations of a variety of actual social groups, and as an assessment of appropriateness in many such instances. Certain familiar types of groups – such as promotions committees, trusts, commissions of inquiry – are assembled precisely in order to make decisions on the basis of specified criteria. Surely groups of this kind can fall victim to the Doctrinal Paradox?

It is worth recalling here that what is required for the Doctrinal Paradox to arise is a relation of strict deductive inference from issues to outcomes, mediated by rules or principles. In the original example, the issues (‘Was there a contract?’, ‘Was there a material breach?’) *dictate*, or in Kornhauser and Sager’s terminology, *are constitutive of*, the outcome.³¹ So in deciding whether the types of ‘everyday’ groups mentioned above are susceptible to the paradox we must imagine them as being beholden to a similarly strict, codified system of rules, not

³¹ This way of modelling the rationality of legal adjudication on logical relationships between propositions represents a substantive position in the philosophy of law which at this point I am assuming only for the sake of argument. In Section 3 this position is called into question, but in this section my point will be that *even if* the norms guiding collegial legal adjudication are rigid and determinative in the way implied by Kornhauser and Sager, we need not think of group judgement and decision-making outside of the legal context as being subject to a similar body of norms.

merely that certain factors are relevant, or are privileged, within the group. If the individuals comprising these groups enjoy leeway – if the rules are only ‘guidelines’, or if ‘outside’ factors (factors not incorporated in the system of rules) can plausibly enter into their determinations – then the paradox will not arise.

To see how this requirement of strictness undermines Pettit’s first generalising move, consider his example of an ordinary group deciding on something which ‘rationally connects’ with other issues. The employees of a company are deciding whether to sacrifice an increase in pay and use the funds saved to jointly purchase a device to guard against electrocution. Now Pettit asks us to ‘suppose for convenience’ that the decision must be made – that is, that each individual must make *their* decision on whether to forgo the increase – on the basis of precisely three considerations: (1) whether the risk of electrocution is serious, (2) whether the device is an effective guard against electrocution and (3) whether the loss of the pay is bearable. A strict rule guides their decision-making: if and only if you answer ‘yes’ to each consideration, must you then decide to sacrifice the increase (and purchase the anti-electrocution device). With this supposition in place, Pettit comes up with a paradoxical profile (Pettit 2003: 171):

Employee	Serious Danger?	Effective measure?	Bearable loss?	Pay-sacrifice?
A	Yes	No	Yes	No
B	No	Yes	Yes	No
C	Yes	Yes	No	No
Majority	Yes	Yes	Yes	No

Of course, when Pettit’s convenient supposition is granted, a case like this will turn out to be paradox-apt. That is because he builds into the case exactly what the generalising move was meant to go beyond: doctrine. This takes the form of the pre-agreed injunction upon employees to decide on the pay-sacrifice on the basis only of the three privileged considerations (1-3). Two features in particular are worth noticing about this pre-agreed principle: it is logically strict, and it constrains the court’s decision-making externally, guiding the decision, but not itself on the agenda in the way the other issues are. These features,

which I think serve to qualify the principle as ‘doctrinal’, and which give the case as a whole its peculiarly juridical nature, cannot be excused as ‘stylised’ features of the hypothetical scenario. As I suggest below, they are necessary pre-conditions of the paradox, and, while some such doctrine may, as a matter of fact, function in the machinations of various kinds of group judgement, there is no reason to think that this is generally required or generally appropriate.

Consider an employee who does not value her own life or the lives of her colleagues, or one who does but who additionally believes that eschatological rapture will commence before the device can be purchased. Less fancifully, think of one who believes simply that the device being considered is not as effective (or perhaps not as cost-effective) as some other device. Such individuals may have answered ‘yes’ to each of the privileged considerations and may even have accepted the principle which privileges them, and yet, on account of their outlying (auxiliary) values and beliefs, may nonetheless not agree to the pay sacrifice. Their thinking will show up as recalcitrant only when the way in which the privileged issues ‘rationally connect’ with the decision at hand is articulated in terms of a strict principle that must be applied via classical *monotonic* logic. But this seems unreasonable, since, as a number of theorists have argued (see, e.g., Antonelli 2005; Ginsberg 1987), perfectly good ‘everyday’ reasoning is *non-monotonic*: “the conclusions drawn may be withdrawn in light of further information, even though all the original premises are retained” (Frankish 2005). From learning that Tweety is a bird, we infer that Tweety can fly, but, on learning further that Tweety is a penguin, we revoke our initial conclusion, that Tweety can fly (but not our premise: we still think Tweety is a bird). So if ‘everyday’ reasoning is – and is appropriately – non-monotonic, then it is not clear that Doctrinal Paradoxes can and should arise in the ‘everyday’ scenarios envisaged by Pettit.

A similar point can be made by considering two further paradox-thwarting employees. The first seeks to attain some personal utility (it could be annoying his colleagues, getting to heaven, whatever) by judging ‘yes’ on each issue but ‘no’ to the outcome; the second lacks certainty on the issue of whether the device will be effective, but, in light of thinking the loss is easily bearable and the danger extremely serious, nonetheless wants to purchase the device. These individuals have not just allowed extraneous factors to enter into their

deliberations, they have allowed *pragmatic* considerations to infiltrate their process of forming *theoretical* judgement. According to a standard, Evidentialist model of theoretical rationality, their thinking must therefore once again show up as recalcitrant. However, it is just these sorts of cases which reveal the inadequacy of the Evidentialist model. If he succeeds in gaining his personal utility, our Pascal-like first employee seems eminently reasonable. (Evidentialists will reply that this is practical reasonableness, and that his 'judgement' is closer to action than to belief – but it is not clear that this helps Pettit). Similarly, his perception of the potential practical importance of the outcome and the *degree* of his beliefs that the loss is bearable and the danger serious ensure that the second (James-like) employee is perfectly justified in drawing a conclusion without having a definitive judgement on whether the device is effective. We may express this by saying that it is surely rational, sometimes, to *try* things – in this case, to *try out* the anti-electrocution device – even when we aren't sure that they will work. As before, the possible intrusion of pragmatic factors into the reasoning of everyday individuals serves to loosen the grip of Pettit's Discursive Dilemma.

We should notice too that the injunction to opt for the pay-sacrifice if and only if you judge the danger serious, the device effective, and the loss bearable occupies an elevated position in Pettit's scenario, reflected by the fact that it is not itself something upon which individual employees get to judge.³² The injunction thus cannot be understood merely as a bi-conditional statement, something which the employees might believe or not – as if a fifth column reflecting individual judgements on the principle

(Serious Danger & Effective Device & Bearable Loss) ⇔ Forgo Increase

could be inserted into the tabulated profile. Instead, the role played by this principle in connecting the group's decision to its reasons is better understood in terms of the notion of a *rule*. It does not just *constrain* but *guides* the group's decision-making, and it *constitutes*, rather than merely *contributes towards*, the nature of the decision at hand. Of course an employee might reject the principle (just as a judge might reject legal doctrine), but there is

³² More carefully, the individuals do not get to judge on this principle *from within* the case at hand. Pettit suggests that we might imagine the injunction arising from a 'prior resolution' (Pettit 2003: 170), but it is interesting to note that he does not consider whether the process through which this resolution might arise is not itself susceptible to some version of the judgement aggregation paradox.

an important sense in which doing so is not a way of participating in the process of group judgement, but is, instead, removing oneself from the decision-making process.

The employees in Pettit's workplace example thus seem to be in the same position which judges in adjudicatory settings appear to occupy, in that they are beholden to strict guiding principles.³³ Naturally, we can imagine a group of people in some extra-judicial context forming a group judgement in this way, but this imaginative exercise has little tendency to indicate a more general paradox if the only way it can be performed is to imagine them as being, as it were, judges without the wigs. On account of its strict, monotonic, and rule-like character, the role of the principle connecting issues to outcome in Pettit's workplace example mirrors precisely the function played, in the original jurisprudential example, of legal doctrine. But this is what Pettit wanted to move beyond, in claiming that the paradox can arise "whenever a group of people discourse together with a view to forming an opinion on a certain matter that *rationally connects*, by the lights of all concerned, with other issues" (2003: 172, italics added). Only if we conceive of the rationality of those connections in terms of a certain kind of logical arrangement ('if and only if') and the nature of those connections in terms of a certain kind of guiding principle or constitutive rule ('on the basis of') will this sort of scenario be generally paradox-apt. If we can reject those suppositions – and, as mentioned, I think Pettit fairly clearly does not wish to rely upon them – then Pettit's first generalising move, the 'social generalisation', is not compelling.

2.2 The Diachronic Generalisation

Pettit acknowledges that this first move does not extend the Doctrinal Paradox as far as he would like, but he seems to consider this to be a matter of limited applicability in the descriptive sense, arising out of the failure of actual groups to achieve a certain kind of ideal:

"The paradox may still seem unlikely to figure much in ordinary social life [because] whereas the judges in a courtroom routinely have to make their judgement by reference to shared considerations, people in other social groups will often reach collective decisions on an incompletely theorised basis... The parties will each vote [for] reasons of their own – reasons related to their own interests or their own judgements

³³ This view of the position of judges is challenged in the final section of this essay; at this point it is assumed for the sake of argument.

of the common interest – and there will only be a partial overlap between the different considerations they each take into account” (Pettit 2003: 172)

Nevertheless, though they may make decisions on an ‘incompletely theorised basis’, social groups that make decisions over the course of time will inevitably have to confront a *diachronic* manifestation of the dilemma:

“Sooner or later such a group is bound to face an issue such that how it should judge on that issue is determined by judgements it previously endorsed” (Pettit 2003: 173)

Whereas groups that act on the basis of settled, shared principles are required to ensure *synchronic* consistency – ensure no contradiction within a single decision between the issues, the principles and the outcome – groups that lack such principles are still required to act consistently over time. Past decisions become the relevant factors, determining what must be decided in the present.

The trouble with this is familiar enough. Even if every individual in the group realises the relevance of past decisions to the current decision, it is hard to see how this history can play a strict determinative role unless the way in which the past is relevant to the present is also, somehow, specified (and there are notorious problems with every attempt to be explicit about this relation). For there will be many possible interpretations of the history of past decisions – many possible ways, to paraphrase Wittgenstein, of going on in the same way – and, since shared guiding principles are *ex hypothesi* not available in these sorts of cases, it is unclear what decides amongst this plethora of interpretations. Another way of putting the point is to say that even when there is some clear requirement that a group’s present judgement must be faithful to its previous judgements, it is not obvious that past judgements are on their own sufficient to determine present cases. Pettit regularly lapses into talk of past decisions ‘dictating’ the outcome of present cases, and of the risk of ‘inconsistency’ between past and present judgement, but, in the absence of shared principles mediating between past and present, it is hard to see how these terms can be meant in the strict (as opposed to the lay) sense. However, as we saw above, the strict sense is precisely what is required for the paradox to arise.

This too can be illustrated by considering one of Pettit's own examples. Pettit (2003: 177-8) asks us to imagine a political party which decides and promises at one point (March) that it will not increase taxes should it be voted into government. Later, after being voted in, it decides (June) to increase spending on national security. Now suppose that, having increased spending on national security, there is no way to increase government spending on anything else without increasing taxes. Yet later still (September), an issue of increased government spending infelicitously arises – say, an issue relating to national healthcare. Pettit thinks that, on pain of inconsistency, which 'would make it a laughing stock among its followers and the electorate at large' (Pettit 2003: 178), the government must act in line with their original promise. They needn't deliberate or vote about this healthcare issue, they must reject any proposal that involves increased spending *simply by applying modus ponens*.

This seems to make the party a hostage to its past decisions in an unattractive – and, arguably, irrational – manner. Consider how we would feel if the healthcare proposal was designed to forestall the otherwise inevitable outbreak of a deadly pandemic. A temporary increase in taxes would then seem quite appropriate, and would certainly not make the party a laughing stock (unless the populace wishes to 'die laughing'). Would this increase mean that the party had changed its mind on the issue of taxes, and broken its promise? Not if we once again allow that the reasoning in this case should be non-monotonic and sensitive to pragmatic concerns. The promise made in March not to increase taxes was made without foreknowledge of the impending pandemic, and in this sense it was always a defeasible promise. It is not, then, contravened by the increase of taxes prompted by the extraordinary circumstances which followed. To put this in terms of the rule-following considerations cited above, the increase in taxes does not mark a failure of consistency (a failure to 'go on in the same way'), because, when the party's promise was originally made, it was not made to bear on the circumstances of an impending pandemic.

So far I have been assuming, with Pettit, that all social groups ought to strive for consistency: that they should always try to go on in the same way. This is because I have been seeking, *contra* Pettit, to voice a familiar scepticism about what such consistency means. But perhaps even this assumption is too strong. In other words, though the courts may be bound by the principle of *stare decisis* – the legal injunction to 'stand by what is decided' – it is not obvious

that we should make similar demands on the sorts of everyday groups with which Pettit is primarily concerned. Here we might wish to take cognisance of an important sense in which the diachronic demands of rationality are usually seen as less severe than the synchronic demands. While we struggle to make any sense of a rational entity asserting a contradiction (claiming, for example, 'B and not-B'), we have nothing like the same trouble comprehending such an entity asserting contradictory statements over time (say, asserting 'B' today, and tomorrow, 'not-B'). This is called 'changing one's mind', and, while there may be good reasons why legal tradition cannot be seen as having 'changed its mind', is it not obvious why these same demands should be placed on ordinary social groups, just as they do not appear, in common discourse, to be placed upon individuals.

Perhaps some 'consistency' over time is required of the court such that how it should decide now is determined in some suitably strong sense by how it decided in the past. This would mean that there might be a plausible diachronic version of the Doctrinal Paradox, but the considerations adduced here are intended to undermine Pettit's claim that there can be a correlative diachronic version of the Discursive Dilemma. It is not clear why we should conceive of regular social groups as being *bound* by their previous decisions in anything like the strong sense in which this might be demanded of judges in the courtroom. Thus Pettit's 'diachronic generalisation' is not compelling.

2.3 The *Modus Tollens* Generalisation

Where the social and diachronic generalisations were concerned to show that the difficulty facing multi-judge tribunals can be instantiated in a broad array of contexts (i.e., that it is a fully *discursive* not only a *doctrinal* difficulty), Pettit's final move targets Kornhauser and Sager's characterisation of the difficulty as a 'paradox'. Pettit contends that casting the difficulty as a 'dilemma' better captures an important logical (rather than contextual) sense in which it is more general than is reflected in the jurisprudential literature.³⁴ Whether 'dilemma' is a more apt term than 'paradox' is not, however, the pertinent issue; what matters is whether the problem really is a more general *sort of problem* than Kornhauser and Sager recognised.

³⁴ List and Pettit's (2001) general impossibility theorem can perhaps be read as the formal expression of this final generalising move.

Recall that Kornhauser and Sager portray the court in paradoxical cases as facing a choice between just two voting protocols – issue-by-issue, and case-by-case. But, as Pettit points out, from a purely logical perspective, these do not exhaust the available options. First, the court could tally all the verdicts, including each individual's assent to the relevant doctrinal principle which guides inferences from issues to outcomes. In that case there will be, according to Pettit, perfect 'individual responsiveness' – every individual's judgement on each issue will be reflected at group level – but at the group level there will be a collective inconsistency. If the method of aggregation proceeds along majoritarian lines then the court will collectively hold:

1. There was a contract
2. There was a material breach
3. If there was a contract and a material breach then the defendant is liable
4. The defendant is not liable

The alternative to individual responsiveness is what Pettit calls 'collective rationality' – the group as a whole seeks to avoid inconsistency by ignoring the views of individuals on one or more of the propositions (1-4). But which ones? It is here that Pettit sees greater logical manoeuvrability than the legal scholars identified:

“another equally salient possibility, neglected as irrelevant in the legal context, is to ignore the majority vote on one of the premises, letting the majority votes on the other premises together with the majority vote on the conclusion dictate the collective view to be taken on that premise” (Pettit 2003: 174)

The logical rule upon which this possibility depends is, of course, *modus tollens*, and Pettit is quite right in saying that deploying it – by ignoring one column of 'premise' verdicts – is another way of attaining collective consistency. In the breach of contract case, for example, the court, having satisfied itself (by majority voting) that the defendant is not liable, might choose, for the sake of consistency, to ignore individual judgements on the question of whether or not a contract existed. I think this can be granted, but we should ask, why does Pettit stop there? There are further ways the court can attain collective consistency. It might, for instance, ignore legal doctrine entirely (and either stop there, or, for the sake of coherence, invent some new doctrine according to which the majority-favoured outcome is

deducible from the majority-favoured issue judgements). Or it might ignore all the judgements made by individual judges (and perhaps pick an outcome from a hat). These are certainly not 'equally salient' possibilities but the crucial point is that logic alone cannot reveal to us their inadequacies. This means that something more than the avoidance of inconsistency must be guiding Pettit's circumscription of the salient options. Since Pettit claims this *modus tollens* move is another route to collective *rationality* (rather than just consistency), this 'something more' must be a conception of rationality that encompasses, but is more substantive than, the constraining dictates of logic. But what this conception is, he does not say.

At any rate, once a thicker notion of rationality is countenanced, a reply on behalf of Kornhauser and Sager can quite easily be made: as far as the court is concerned, the *modus tollens* move is by no means comparable to the case-by-case and issue-by-issue protocols. Though it solves the problem of logical inconsistency, ignoring the judges' views on a single issue (say, the issue of whether a contract existed) would simply not be *coherent*. Since the court is charged with arriving at outcomes *on the basis of* issues and doctrine, it is, as we might like to put it (adopting Pettit's preferred logical idiom), precisely in the business of applying *modus ponens*.

In this regard the role of doctrine in legal adjudication is doubly significant. As we have been assuming (along with Pettit, and Kornhauser & Sager), doctrine is not itself something upon which judges get to cast judgement, at least not in the process of judging cases. Pettit seems conscious of this, since his *modus tollens* manoeuvre extends only to the possibility of ignoring the views of judges on some issue, and not to ignoring the legal principles connecting issues to outcomes. Legal doctrine is thus insulated from the arrow of *modus tollens*, but a proper appreciation of the role of legal doctrine would, I think, equally insulate the columns of issue verdicts which Pettit's proposal wishes to target. This is because the function of legal doctrine is to articulate the constitution of the case before the court – to say which issues are relevant, and how. In the breach of contract example, the relevant legal doctrine specifies the way in which a liable verdict is constituted by the issue of contract and the issue of breach. To think that ignoring the views of the court on one of these issues could nonetheless be a way of remaining faithful to doctrine (merely because doctrine itself has not been abandoned or

‘falsified’ by *modus tollens*) is to radically misconceive the nature of legal adjudication. Ignoring the views of the court on a particular issue is itself a way of rejecting legal doctrine, and, since rejecting legal doctrine is (as Pettit tacitly acknowledges) *not* a plausible way of preserving collective consistency, we must conclude that there is, after all, no *modus tollens* manoeuvre available to the court.

3. Conclusion: Reason and Law

Taken together, the foregoing objections amount to a rejection of Pettit’s generalising move from the Doctrinal Paradox to the Discursive Dilemma. Neither the social nor the diachronic generalisation succeed in showing how the difficulty identified by Kornhauser and Sager can appropriately arise in extra-judicial contexts without the mediating presence of doctrine. Moreover, since it is the special nature of the judicial context which reveals the incoherence of certain logically possible evasive manoeuvres, it is not clear why we should view the difficulty as amounting to a grand choice between individual responsiveness and collective rationality, rather than a troublesome, protocol-related kink in the legal system.

In making the argument in this way I have been prepared to assume – with Pettit *and* his opponents, Kornhauser and Sager – that legal reasoning is ‘doctrinal’ in the way required for the Doctrinal Paradox to arise. This has meant that I have sought only to block the claim that reasoning outside of the courtroom should be viewed in similar, doctrinal terms. But to end off I would like to consider another way my argument might have gone – an alternative which seeks to affirm a link between the juridical and the more broadly discursive while suggesting that neither can properly be understood in doctrinal terms. This alternative is worth considering because it supplements the negative case against Pettit with something more positive. In particular, it follows through a little more on the insinuation with which I began, that the presence of the conditions that give rise to the Doctrinal Paradox amount to a subjugation rather than an exemplification of reason.

3.1 Legal Realism and Legal Formalism

We can begin by recalling the original jurisprudential example of the Doctrinal Paradox and asking after the *form* of the relevant legal doctrine (the injunction to judge liable if and only if you judge that there was both a contract and a material breach). The precise form of legal

doctrine will of course depend on the sort of legal tradition in which it features: 'common law' systems give authority to precedent (past cases), while 'civil law' systems are based on general legal principles or statutes. But I think we can abstract from these differences here, since we shall not wish to confine our conclusions to only one legal framework, and anyway the two systems are not quite as diametrically opposed as is sometimes thought. Even common law is not case law all the way down (there must at least be a general principle of *stare decisis*), and, likewise, civil law cannot consist only of statutes (it must be grounded in some practice of interpreting and applying statutes, otherwise an endless regress of statutes would be needed). So let us instead reflect on two dominant conceptions of legal doctrine in the context of the judiciary – 'legal formalism' and 'legal realism'.

According to *legal formalism* (see, e.g., Scalia 1989) what judges are doing when they discharge their duties is bringing some already-existent law to bear on the case before them. Their job is to thus to know what the law is, and to identify which part of it is relevant, before applying the legal doctrine identified as relevant to the case at hand. Though these tasks require significant intelligence, experience, and disgression, this conception accords judges only a derived sort of authority. What is genuinely authoritative is the law, and the duty of judges is to be *responsible* to its authority.

The alternative conception of legal doctrine in traditional jurisprudence is supplied by various forms of *legal realism* (see, e.g., Holmes 1897). According to this conception, legal doctrine is not something to which judges are beholden, but rather something which they themselves forge through their judgements – law is 'judge-made'. This is a complete reversal of formalism: judges are not seen to be responsible to an already-existent standard but are instead the authoritative creators of law. But this leads to certain scepticism about the content of law, bordering on legal nihilism. To say, for instance, that such-and-such is unlawful is to make "a matter-of-factual *prediction* about what a judge would decide" (Brandom 2012: 19), and from here it seems only a short step to the realist adage that the law is founded on 'what the judge ate for breakfast'.

With these rough sketches in hand let us return to the Doctrinal Paradox, where it seems that the formalist conception is presupposed. For if judges are not answerable to some pre-

existing standard, but are instead *law-makers*, as the realist has it, then it is hard to see how there could be any meaningful problem with the set of views arrived at by majority voting on all issues. The problem only arises when those views are set against some pre-existing legal doctrine with which they are incompatible, so if the derivation of verdicts is itself the process by which legal doctrine is determined then no such incompatibility could ever arise.

The Doctrinal Paradox thus seems to beg the question against legal realism, but, since legal realism has the above-noted sceptical/nihilist implications, this is not the grounds of my complaint. Instead, the criticism I wish to level against the promulgators of the Doctrinal Paradox and its generalisations is that they have also begged the question against another, intermediate (and thus less easily classifiable) conception of legal doctrine. This conception depicts judges as both responsible to and authoritative over legal doctrine, in a way which can be seen to exemplify the structure of reason more generally. The picture of reason which emerges from this exemplification is one which recognises both the autonomy and the genuine boundedness of reasoners, and it is reason thus exemplified which I think Pettit misses out on.

3.2 Law as Interpretation

Ronald Dworkin's (1986) conception of 'law as integrity' is perhaps the best expression within mainstream legal philosophy of the position which lies between legal realism and legal formalism:

"Law as integrity deplores the mechanism of the older "law is law" view [i.e., legal formalism] as well as the cynicism of the newer "realism". It sees both views as rooted in the same false dichotomy of finding and inventing law" (Dworkin 1986: 228)

How do we avoid this dichotomy? Dworkin contends that we avoid it by recognising that adjudication is an essentially *interpretive* enterprise:

"legal claims are interpretive judgements and therefore combine backward- and forward-looking elements; they interpret contemporary legal practice seen as an unfolding political narrative" (Dworkin 1986: 225)

Dworkin illustrates this by drawing an analogy between legal adjudication and the writing of a chain novel. In a chain novel, successive authors have the task of contributing successive chapters to the novel and each seeks to develop the overall work while still remaining faithful to trajectory he perceives as having been forged in previous chapters. In this way the contributors must be both 'backward- and forward-looking'. They must identify certain features of the extant work as *salient* and seek to build on these in their own contribution, all the while recognising that this amounts to an interpretation of the unfolding novel which future contributors may ignore, reject or emend. In the same way, claims Dworkin (1986: 229), "Judges... are authors as well as critics. A judge... adds to the tradition he interprets; future judges confront a tradition that includes what he has done."

In a recent paper, Robert Brandom (2012) has sought to frame this Dworkinian idea in terms of the broadly Hegelian approach to reason and discursiveness which he has been expounding for some time (see esp. Brandom 2009). The result – Brandom's Hegel's Dworkin – is a novel contribution to the philosophy of law, and one which should hold particular interest for those who, like Pettit, think the Doctrinal Paradox has broad import. Brandom suggests that we read Dworkin as gesturing at a radical reconception of the way in which legal concepts are 'determinately contentful'. Take the concept of 'tort-liability' which is the central legal concept in the standard example of the Doctrinal Paradox. The finding/inventing dichotomy rejected by Dworkin sees the content of this concept either as always already settled (formalism) or as always awaiting settlement (realism). Either way, judges are 'determining' the law, since 'determine' can mean either *find* to be the case, or *make* to be the case. But the notion of determinate-ness favoured by Brandom (which he identifies in Hegel, and confers on Dworkin) is one which rejects the thought that we have to *choose* between these two senses of 'determine'. When judges determine the content of tort-liability they are both finding it and making it, and so the determinate content of the concept is dynamic, not static.

If legal doctrine is construed in these interpretive, dynamic terms, then the grip of the Doctrinal Paradox is severely loosened. For the pressure to avoid certain sets of group-level judgements is only intelligible against the backdrop of a conception of legal doctrine which sees its content and relevance to the case at hand as already-settled. If, by contrast, judges in paradoxical scenarios are seen as involved in the sort of enterprise envisaged by Dworkin and

Brandom – in which the law is seen as continually being determined (being both found and made) – then there is no settled standard against which inconsistency or even incoherence can be decisively revealed. But this does not mean that judges are radically unconstrained in the manner imagined by extreme legal realism. The interpretive approach sees judges as responsible to the legal traditions in which they are located, but authoritative over their interpretations of these traditions.

Apart from mandating a scepticism about the force of the Doctrinal Paradox, the interpretive view of legal adjudication can also serve as a model for an alternative general conception of rationality. As we have seen, Pettit tends to understand reason itself in what I have been calling ‘doctrinal’ terms: he views ordinary discursive creatures as *bound* by already-settled norms or principles, which they must obey on pain of irrationality. In this way Pettit sees only the responsibility of reasoners and not their authority, and hence the formalist model of legal adjudication implicit in the jurisprudence of Kornhauser and Sager appears to him as a paragon of the exercise of reason itself. But pursuing the interpretive line, as Brandom (2002, 2009) has done, can take us in precisely the opposite direction: legal adjudication can then be seen to exemplify reason itself by illustrating some of the ways in which discursive creatures like us are bound by rational norms over which we ourselves have some authority. The relations of reciprocal authority and responsibility borne by judges towards their predecessors and their successors in a legal tradition, and through which the dynamic content of legal concepts is continually being *determined*, provide a model of the way our general discursive practices are fundamentally both social and historical. Just as Pettit’s doctrinal conception of reason ends up having broad social ontological consequences, so too might Brandom’s Hegelian conception of reason. Just what these consequences might be strikes me as a worthwhile topic for future research.

Essay 3: Being and Becoming in the Theory of Group Agency

Introduction

Philip Pettit's theory of group agency is primarily concerned with the reasons for counting some social groups as rational agents over and above their members, and what the autonomy of such group agents consists in. But though he provides an admirably clear, detailed account of what *being* a group agent involves, Pettit has paid rather less attention to the question of how group agents come about – how they *become* the group agents they are. Here I wish to explore this somewhat neglected question about the emergence of group agents.

One of Pettit's key claims about the emergence of collective agents or corporate persons is that they "made, not born" (Pettit 2007: 495), and this raises the question of just who it is that makes them. His original treatment of the Doctrinal Paradox (Pettit 2002; List & Pettit 2001) and subsequent forays into organisational design (esp. List & Pettit 2011) suggest that the collective itself somehow 'gets its act together', transforming itself from a set of individuals into an autonomous group agent.³⁵ A group agent is thus, for Pettit, a sort of 'self-made man', and this introduces a particular kind of bootstrapping puzzle about the process through which such agents emerge. Specifically, it is hard to see just how the step that Pettit thinks collectives must take in order to *become* agents – the 'collectivisation of rationality' – can be taken without the collective in question already *being* an agent.

My aim in this essay is to see whether Pettit succeeds in providing a non-circular account of the process whereby a collective incorporates into an agent or person.³⁶ In Section 1, I explore what Pettit considers to be the first stage of this process, in which a collective incurs a certain kind of *pressure* – the pressure to incorporate, to become a group agent. For Pettit, this stage of the process is exemplified by the much-discussed Doctrinal Paradox of judgement aggregation, and so the question I pursue in this first section is whether the paradox can have any traction for a collective that is not already incorporated. In Section 2, I

³⁵ For the sake of clarity in this essay I use the term 'collective' to denote an un-incorporated but potential group agent, and 'group' for an already incorporated group agent.

³⁶ Worries over the possible circularity of accounts of collective intentionality have been explored before, notably by Tollefsen (2002b) who targets the views of Tuomela (1995) and Gilbert (1996). As far as I know Pettit's theory of group agency has not been subjected to similar critique.

turn to the second stage of incorporation, in which the collective responds to the pressure to incorporate by undertaking to exercise reason at the collective level. Pettit thinks there are a variety of decision-making procedures which might accomplish this 'collectivisation of reason', and the question I ask about these is whether they too might actually presuppose an already-incorporated group agent. In the end I think that Pettit fails to dissipate the bootstrapping problem bruited above, and, in Section 3, I diagnose this failure as the product of a tendency to overlook the role that inter-subjective attitudes in a broader discursive community might play in instituting a group's status as an intentional subject or corporate person.

1. The Pressure to Incorporate

The lesson of the Doctrinal Paradox, and Pettit and List's (2001) famed impossibility theorem, is that two requirements we may wish to place on the formation of collective attitudes are in fact in tension with each other. The first is that the collectively-held attitudes ought to be sensitive to the attitudes of the collective's members; the second is that the attitudes collectively endorsed should comprise a set that would be rational to hold. Now Pettit's key claim is that certain kinds of collectives will face an overwhelming pressure to give priority to this second requirement, even though doing this may lead them to violate the first requirement – they may end up espousing views which many (even all) of their members individually reject. To make sense of Pettit's account of how group agents emerge we need to make sense of this 'pressure' to exercise reason at the collective level.

1.1 Paradox and Pseudo-Paradox

The Doctrinal Paradox is standardly illustrated with an example involving someone being sued for damages arising from an alleged breach of a contract (see pp. 36-7 for this example in full). Legal doctrine dictates that the decision of the court – as to whether or not the defendant is liable – must depend on just two issues: (1) whether a contract existed prohibiting the defendant from performing a certain action, and (2) whether or not the defendant performed that action. Paradoxical profiles are those in which the collective views – i.e., those derived from majority voting – are inconsistent with legal doctrine.

Pettit has devoted considerable effort to generalising the scope and significance of the standard renditions of the Doctrinal Paradox – showing how structurally identical problems

can arise for a great variety of collectives in a great variety of circumstances, even without the mediating influence of ‘doctrine’.³⁷ But for all his generalising efforts, Pettit has, to my mind, neglected the task of *circumscribing* the difficulty confronted by collectives in these paradoxical scenarios. For it is possible – and, I think, instructive – to generate profiles which *resemble* paradoxical cases but in which, intuitively, there is no real problem for the collective. As we shall see, reflection on such ‘pseudo-paradoxes’ can make it tempting to think that the pressure to collectivise reason will only be felt by groups which are already incorporated.

Imagine that a television station wishes to know whether its viewers would like the Olympic diving event to replace one of the two programmes – CSI or ER – ordinarily screened between 7pm and 9pm on a certain night. Upon interviewing a (very) small sample audience, the following profile is generated:³⁸

Person	Replace (CSI or ER) with diving?	Drop CSI?	Drop ER?
A	Preferred	Dispreferred	Preferred
B	Preferred	Preferred	Dispreferred
C	Dispreferred	Dispreferred	Dispreferred
Majority	Preferred	Dispreferred	Dispreferred

This profile may well present a problem for the TV station, but it is not obviously problematic for the TV-watchers. A majority of the individuals interviewed preferred that regular programming be interrupted in order for the diving to be screened, but most did not want CSI to be dropped *and* most did not want ER to be dropped. Is ‘the majority’ confused here, even though each individual is quite clear and consistent? No, because the apparent inconsistency involved in ‘what the majority wants’ vanishes when we observe that the majority preference on each issue is constituted by a unique pair of individual preferences. Although it is inconsistent to wish for either CSI or ER to be dropped, but neither CSI nor ER to be dropped, the rational requirement that such inconsistency be avoided only applies to unitary entities. Since the majority featured in each column is, strictly speaking, a *different* majority each time,

³⁷ These efforts are discussed at length in Essay 2.

³⁸ The term ‘dispreferred’ which features in the profile is adopted from List and Pettit (2011) to denote the attitude of being actively opposed (i.e., not simply indifferent) to some state of affairs.

the bottom line of this profile reflects disagreement amongst the people interviewed rather than inconsistency.

Why can we not then say, to return to the standard example of the Doctrinal Paradox, that there is no failure of rationality in the court's verdicts derived from majoritarian voting on all issues? This would mean that the court judges that (a) a contract existed, that (b) a breach occurred, that (c) if a contract existed and a breach occurred then the defendant is liable³⁹, and that (d) the defendant is not liable. This sounds irrational only because we have surreptitiously slipped from talk of what 'a majority of judges' thinks to talk of what 'the court' thinks; if we remember that the view of the court is only an arithmetic function of the individual judges' views then this appearance of irrationality disappears. In other words, to say that there is a failure of rationality appears to beg the question of the court's status as an entity in its own right – one which is subject to certain constraints of rationality.⁴⁰ But if paradox only arises for groups that are already rational entities in their own rights, then it seems it could not provide the impetus for the emergence of such entities.

1.2 Purpose and Paradox

Clearly, Pettit needs a principled way of distinguishing paradox from pseudo-paradox, and a way of deeming the bottom line inconsistency in the judicial profile pernicious without presupposing that the panel of judges is already incorporated. How might this be done? Perhaps by drawing on the idea that certain social collectives are *purposive*.

The sample group of TV-watchers is what Pettit calls a 'mere collection' (List & Pettit 2011: 31-2): they have no common purpose or even any knowledge of one another's contribution to the interview process in which they each participate. By contrast, the court is a 'purposive collectivity' (Pettit 2002, 2003) required to *act* in designated ways – imposing penalties, awarding damages, sentencing, etc – which uphold 'legal doctrine'. Because the court's judgements are 'near the coal-face' (Pettit 2009: 77) of their action they cannot, on pain of jeopardising the court's agential capacities, be inconsistent: "Let an agent try to act on

³⁹ This is based on the assumption that the judges agree on the relevant legal doctrine.

⁴⁰ Such question-begging is, arguably, exhibited in Pettit's tendency to describe *the set of views* which issue—by-issue majoritarian voting would deliver as 'irrational' instead of simply inconsistent (see esp. Pettit 2002, 2003; List & Pettit 2002).

inconsistent representations or motivations... and there will be a straightforward breakdown: actions will be supported that cannot be realised" (List & Pettit 2011: 25).

This distinction between purposive collectivities and mere collections seemingly helps Pettit to make non-circular sense of the pressure to incorporate. A collective might be recognised as purposive – it might have certain goals which it needs to pursue – and because of this, *as a response*, it ensures that its collective views are rational. According to Pettit, standards of rationality are "nothing more or less than desiderata of agency: standards such that agents will generally do better as agents by robustly satisfying them" (Pettit 2007: 497). In order to be able to effectively *act* in fulfilment of its purpose, a purposive collectivity must take steps to ensure that its collectively-held attitudes are, at a minimum, mutually consistent.⁴¹ Thus it is the panel of judges' purposiveness which distinguishes them from the TV-watchers, and which ensures that they find themselves susceptible to the distinctive pressure exemplified by the Doctrinal Paradox. In short, the fulfilment of purpose requires agency, and agency requires rationality – where, crucially, what is 'required' is not pre-required but can be provided responsively.

One might still wish to press the original bootstrapping objection here, by claiming that identifying something as a purposive collectivity must presuppose that it is a rational agent, since a purpose could not be fulfilled by a collective that was not already a rational agent. But this line of objection is mistaken: it elides the distinction between potentiality and actuality, overlooking the fact that treating something as a purposive collectivity can itself serve to instil or awaken the very rational-agential capacities required for the fulfilment of its purposes. This kind of bootstrapping, according to which certain capacities are *brought about* by being taken to be *already there*, is often thought to underpin the developmental processes in individuals. For example, some theorists (e.g., McGeer 2002; Pettit 1995) have claimed individuals come to be trustworthy only by being trusted – that is, by being treated as though they were already trustworthy. Similarly, it has been claimed that actual fitness to be held responsible for one's actions results from a process of responsabilization (Garland 2001) – of being treated

⁴¹ According to Pettit, to be rational one must form attitudes which the evidence favours ('attitude-to-evidence' rationality), one must not have inconsistencies in one's body of attitudes ('attitude-to-attitude rationality'), and one must tend to act for the realisation of one's preferences in line with one's beliefs ('attitude-to-action rationality').

as though one were fit to be held to responsible. Such bootstrapping seems to serve a powerful purpose in the individual case and it should not be ruled out in the case of social groupings.

A better line of objection targets the purported link between purposiveness and the requirement of agency. An agent, according to the conventional picture Pettit adopts, is something with beliefs and desires, and the capacity to intervene in the world for the fulfilment of its desires in line with its beliefs (Pettit 2002, 2007, 2009; List & Pettit 2011). But collectives are composed of individuals who are themselves agents, so why can't the individuals comprising the collective co-ordinate their actions so as to serve the purposes of the collective? Purposiveness may require agency but it is not clear at this point why it requires the collective's own agency.

To answer these questions we need to attend more carefully to the notion of purpose involved in the idea of a purposive collectivity. Once again we may find it instructive to reflect on the differences between the TV-watchers and the panel of judges. Why are the TV-watchers not counted as purposive? Certainly they are interviewed for a purpose. The crucial difference between them and the panel of judges seems to be that it is not *their* purpose but the TV-station's which their stated preferences are rung in to serve. (To see this, imagine that the interviewees pooled their resources and bought the TV-station; then the pattern of their judgements on the scheduling issue would be *their* problem, and the profile would be paradoxical.) By contrast, the panel of judges seems to have a purpose of its own – something which the individuals on the panel collectively pursue – so perhaps it is this proprietary purpose which prompts certain collectives to form their own centres of agency.

What then does it mean for a collective to have a purpose of its own? Presumably it means that there exists some goal that the collective strives to achieve or some role that it seeks to perform. But one thing it cannot mean, if Pettit's account of the emergence of group agents is to be coherent, is that the collective has a kind of *attitude* of its own – something like a group intention. For if the purpose of a purposive collectivity were tied to a group-level intention then it would seem that we would already have an incorporated group agent – something

which is already an intentional subject in its own right – and so the having of a collective purpose could not be integral to the process of becoming a group agent.

Pettit sidesteps this concern by appropriating a key concept from the literature on collective intentionality – the concept of ‘joint intention’ (Bratman 1999; Tuomela 1995; Gilbert 2001). According to Pettit, a joint intention can be distinguished from a group intention since it is a configuration only of individual attitudes, such that (List & Pettit 2011: 33):

- (1) each of the individuals in the collective intends that they together promote some purpose,
- (2) each of the individuals intend to play their part in fulfilling that purpose,
- (3) each of the individuals forms these intentions (1&2) because they believe others have formed similar intentions, and,
- (4) this is all a matter of common awareness.

The notion of joint intention appears to provide Pettit with a route out of the bootstrapping worry, at least insofar as the worry concerns the pressure to collectivise reason. The first clause (1) tells us that, where there is joint intention, each of the members of a collective intend that they jointly serve some purpose. This means that each of the members has an individual interest in taking steps to preclude the sorts of inconsistencies in collective-level judgements which can (as the Doctrinal Paradox teaches us) result from perfect individual responsiveness. This is because, as Pettit tells us, inconsistencies cannot provide a sound basis for the fulfilment of the collective’s purpose. The pressure felt by certain collectives to collectivise reason is precisely the realisation that their purposes cannot be served if inconsistencies in group-level judgements are allowed to arise. But when commitment to these purposes is understood in terms of joint intention it is an entirely individual affair – there is no group intentional subject presupposed.

2. The Response to the Pressure

Perhaps then a conceptual wedge can be driven between the notion of members of a collective severally sharing a purpose and the notion of a group with its own purpose. If so, then the fact that members of a collective share a purpose does not imply that the collective is already incorporated – and hence it is possible that its purposiveness, understood in terms

of joint intention, might call for incorporation without presupposing it. This would mean that we can make sense of the pressure to incorporate in a way that does not make it unnecessary (i.e., such that the only possible subjects of that pressure are already incorporated). But this does not yet mean that the coherence of Pettit's account of the emergence of group agents is secured, because it is not clear at this point how a collective could possibly *respond* to that pressure without already being a group agent.

We might worry that what is required as a response to the pressure exemplified by the Doctrinal Paradox is the collective's own response, where this is understood as some sort of group action. But such a response could not of course be part of the account of the emergence of group agency, since it would quite clearly require the collective to already be an agent. So what is needed is a member-level story – one which doesn't pre-require but might still be able to institute a group agent. Pettit seeks to provide such a story by endorsing a second notion prevalent within the collective intentionality literature – the notion of *joint action*, which he distinguishes from group agency.

Pettit claims that several individuals can co-participate in a joint action on the basis of their joint intention without constituting a group agent (Pettit & Schweikard 2006: 19). But although it need not, joint action can, according to Pettit, give rise to group agents: a collection of individuals can jointly act in certain ways so as to transform themselves into a group agent (Pettit & Schweikard 2006: 33). This is the second, response-stage stage in Pettit's account of how group agents come into existence, and its tenability seems to depend on two things, which I will consider in turn. First, it depends on whether it is plausible to agree with Pettit that joint action does not imply an already-incorporated group agent. Second, it depends on whether the steps that need to be taken to transform a collection of individuals into a group agent are indeed candidates for joint action.

2.1 Joint Action versus Group Agency

According to Pettit (Pettit & Schweikard 2006: 23-24), joint action is action performed by members of a collective on the basis of their joint intention, where this is understood in terms

of the conditions (1-4) discussed above.⁴² As noted, these conditions seem to relate only to individual attitudes, so it would appear that the joint intention upon which joint action is based does not presuppose a group intentional subject. However, we may wish to inquire after the nature of the intention mentioned in the first condition, which says that each of the individuals in the group *intends that* they together promote some purpose.

What does it mean to *intend that* X? We have, I take it, a reasonable grasp of what it means to *intend to* X – this is an attitude an agent takes up towards a possible future action of her own, something the agent can herself perform. But accounts of collective agency face a distinctive difficulty when it comes to the matter of intention, and this has led to the introduction of the notion of intention-that. The problem, in short, is that one can only *intend to* do something which one has the capacity to do, but since no single participant in a joint performance has the capacity to individually produce a joint performance, no single participant can *intend to* perform the joint performance. But, if no-one intends the collective performance then we are seemingly left with an unintended performance, and hence not an action at all. This problem has led some theorists (e.g., Quinton 1975) to conclude that there is no such thing as joint action, and others (e.g., Gilbert 1989) to conclude that any instance of joint action implies a group agent or ‘plural subject’ (i.e., something which can *intend to* perform a joint performance). But the concept of intending-that seems to make room for a third possibility – the possibility of genuine joint action that does not imply a group agent. This is the possibility Pettit wishes to defend, and I think *must* defend if his account of how group agents come into being is to be coherent.

What does it mean for the members of a collective to intend that they together enact a joint performance? The intuitive distinction between intending-to and intending-that is that, whereas the former is an attitude directed upon a course of action, the latter is directed upon a state of affairs, and hence seems to qualify as a kind of ‘propositional attitude’. However, if the special role played by intention in the production of action is to be preserved, then intention-that needs to be sharply distinguished from other propositional ‘pro-attitudes’, such as desiring and wishing. One can desire or wish for states of affairs that are entirely beyond

⁴² Pettit and Schweikard (2006: 24) claim this is closest to Bratman’s (1999) analysis of shared agency.

one's sphere of influence, but it would not make sense to say one can intend these things. (For example, I can certainly desire or wish that it stop raining, but I cannot intend that it stop raining.) So it would appear one can only have an intention-that if the content of one's intention falls within one's sphere of personal influence. But does this not simply lead us back to the original difficulty, which was that a joint action cannot (by definition) be something within any *individual's* sphere of influence? Not necessarily, since Pettit thinks intenders-that need only have *partial* – not necessarily total – control over what they intend: “I intend that something happen only if... I want it to happen and I am in a position to do *something* about it” (Pettit & Schweikard: 23, italics added).

This conception of intention-that does seem to allow for joint action without a group agent. Since each group member has some part to play in a joint action, the individual members can *intend that* they together perform a joint action because they each *intend to* do their part in the collective performance. But even if we grant Pettit this somewhat contested notion it is not clear that the joint action it supports could be the means through which group agency emerges. That is because it is not clear that the steps that would need to be taken for a collective to become a group agent are candidates for joint action. Let us turn to this second question now.

2.2 Collectivising Reason as Joint Action

What steps must be taken for a collective to transform itself into a group agent? We know from Pettit's discussion of the Doctrinal Paradox that he thinks they need to *collectivise rationality*, but is this a potential joint action? More particularly, can each individual member *intend that* they together collectivise rationality? This might seem a peculiar question, especially in light of the fact that we are not here questioning whether collectivising rationality is something that groups can do (our question is about *who* does it, and, specifically, whether it can only be done by a group agent). But if groups do collectivise rationality, and we agree with Pettit and List's (2006, 2011) supervenience thesis – that the attitudes and actions of the group supervene only on the attitudes and actions of individual members – then it seems the collectivisation of rationality must be something which individual members can intend. Intending-that meant taking up a pro-attitude towards a proposition – in this case, that the collective collectivises rationality – and committing to

playing one's part in that proposition being realised. But since, by the supervenience thesis, individuals cannot help but be involved in anything the collective does (including becoming rational at the collective level), it seems that individuals can certainly so intend, and hence that the collectivisation of rationality can be a joint action.

Once again, however, we should be careful, for it is not obvious that simply being a part of something amounts to 'playing one's part'. Consider another example: imagine that, for whatever reason, I wish to discover *completely by accident* that my partner is having an affair (perhaps I wish to separate, but think it is only legitimate for victims of adultery to instigate these things, and I also don't want to snoop or pry). Since discovering completely by accident that one's partner is having an affair is something that could happen, it seems plausible that it is something one could wish for. But is it something one could intend? Certainly one could not *intend to* make this discovery since what is discovered is not wholly within one's personal sphere of influence (i.e., some part of it is in *another's* sphere of influence). Could I *intend that* I accidentally discover my partner's infidelity? I do not think so, even though it is plausibly something I can take up a pro-attitude towards and something, were it to be realised, which would necessarily involve me. The reason I cannot intend it is that the way in which I would have to be involved is not as an agent. Since what has to happen for the wish to be realised has to happen *accidentally*, I cannot intend to actively play any role in bringing it about; playing a role would, precisely, defeat the fulfilment of that wish.

I think that something similar goes on when we try to conceive of the members of a collective intending that they together collectivise rationality. Collectivising rationality, as the Doctrinal Paradox teaches us, is in direct conflict with individual responsiveness. What this means is that, although this collectivisation must necessarily involve individuals, it does not cast them in the sort of role they can intend beforehand to perform. To commit to the collectivisation of rationality is to commit to being less than an individual agent in the exercise of one's group role; in acting and thinking *for* the group one is not thereby performing as an autonomous agent, but as part of the group's subpersonal machinery. When one intends, one takes up an attitude towards a potential future action of one's own (this can be part of a joint action), but what is required in the collectivisation of rationality is not the performance of one's own action, but the performance of someone else's action – the group's (cf. Rovane 2004). This

means that the collectivisation of rationality is not a potential joint action; it is a singular process whereby a new centre of rationality and agency is formed, and this process is not one in which group members can intelligibly intend to participate.

Perhaps Pettit is able to sidestep this concern too, since he does not quite say that collectivising rationality is itself the joint action through which groups emerge. Nonetheless he does believe that joint action can effect the collectivisation of rationality, and so bring about group agents:

“First, the members act jointly to set up certain common goals and to set up a procedure for identifying further goals on later occasions. Second, the members act jointly to set up a body of judgements for rationally guiding action in support of those goals, and procedures for rationally developing those judgements further as occasion demands. And third, they act jointly to identify those who shall act on any occasion in pursuit of the goals” (Pettit & Schweikard 2006: 33)

What is the difference between this and jointly performing the action of collectivising rationality? The key difference seems to lie in the introduction of ‘procedure’, by which is meant more or less explicit rules to guide the internal machinations of the group. If by joint action these procedures can be inculcated, and if action conforming to these procedures will count as collectively rational, then it appears that joint action can, in the end, bring about group agents (though that action is not itself collectivising rationality).

But what will these all-important procedures be? Pettit considers the virtues of a number of functionally explicit organisational procedures, including the ‘sequential priority rule’ in which logically interconnected issues are placed in a sequence and issues only go to a vote so long as they are not already determined by their logical relation to prior, already-decided issues. In the end Pettit thinks that though they guarantee consistency in group-level judgement, the rigidity of functionally explicit procedures such as the sequential priority rule may cause groups endorsing them to violate some other demands of rationality:

“When I realize that some propositions that I believe entail a further proposition, the rational response may well be to reject one of the previously accepted propositions rather than to endorse the proposition entailed. Those are the undisputed lessons of

any coherence-based methodology and the group that operates under a sequential priority rule, or under any variant, will be unable to abide by them" (Pettit 2007: 511)

So if functionally explicit procedures are unable to secure collective rationality, what sort of organisational mechanisms does Pettit think groups should adopt in order to transform themselves into group intentional subjects? To answer this we need to attend to a distinction Pettit makes between 'reason' and 'rationality'. As we saw above, rational standards, for Pettit, are nothing other than the desiderata of agency; rational creatures are those who tend to believe what the evidence favours, whose store of beliefs and preferences are largely consistent, and who tend to act in line with their intentions. They are veracious, consistent and enkratic because this is the best way to be an agent. But none of this requires a rational creature to attend to the propositional contents of her attitudes and form meta-propositional attitudes about them – asking, for example, whether this belief is *true*, whether that preference *implies* another, whether these judgements are mutually *consistent*, and so on. According to Pettit, simple rational agents can get along adequately without being able to attend in this higher-order way to their thought; their rationality – the extent to which they meet the standards of rationality laid out above – is "secured by their make-up or design" (Pettit 2007: 495). But sophisticated rational agents, who have the capacity to form attitudes about their attitudes, can go beyond this, reinforcing their rationality by means of *reason*:

"To be able to reason... is to be able to conduct an intentional activity that is designed – and perhaps explicitly intended – to raise the chance of satisfying [rational] desiderata" (Pettit 2007: 499)

Pettit thinks that the robust rationality of groups will only be secured if the group can succeed in reasoning, and he thinks that there is at least one sort of organisational mechanism which can achieve this: the straw vote. A straw vote is simply a non-binding vote, through which a group's judgement can be registered without being concretised; the views thus reached can always be revised later if the group so decides. A straw vote can be used to temper the inflexibility of the likes of the sequential priority rule, thus allowing groups to be more sensitive to what rationality (not just consistency) requires. For example, a group might vote on each proposition in a sequenced set, and, if their voting generates an inconsistency, they "consider all the different possible ways in which previously formed attitudes or the new

attitude could be revised so as to restore consistency... [and] take a vote... on which of the possible revisions to make” (List & Pettit 2011: 63).

When groups deploy a straw vote procedure they are clearly doing what Pettit would call reasoning. This is because they are concerned with the consistency of the judgements they have defeasibly adjudicated – standing ready, should inconsistency be discovered, to make the appropriate revisions *at any point within the overall body of judgment* (depending on what the evidence favours). In this way the straw vote allows the collective to be more robustly rational, since the collective is required to reflect on the overall rationality of their collective judgements.

Once again, this much can be readily accepted, since the degree to which rationality is or is not attained by a collective is not our primary concern. Our concern is with the process by which collective rationality is attained and, more specifically, with whom it is that institutes and handles that process. The suspicion we have been pursuing is that the process by which rationality is collectivised is a process which pre-requires a group agent, and so cannot be the process through which group agents emerge. This means that the question we need to ask about the sort of collective reasoning implicit in straw vote-type procedures is whether it is best seen as an achievement of the individuals in the collective or of an already-formed group agent.

Pettit’s discussion of the reason/rationality distinction as it pertains to individuals suggests that when that distinction is transposed to the social level, it must be the group itself that reasons:

“The rationality of the simple creature is realised sub-personally so far as there is nothing *the creature* can do in order to improve its rational performance... We reasoning creatures transcend this limitation. We do not have to rely on the processing for which our nature programs in order to be rational. *We* can do something about it... The transcendence that reasoning achieves gives [individuals] a degree of personal control over [their] own rational performance” (Pettit 2007: 502, italics added)

If the exercise of reason in individuals is a personal-level phenomenon it would seem that the exercise of reason in collectives must be a group-level phenomenon – and the control over its rational performance which such exercise strives for is something the group itself achieves. Though individuals participate in the straw vote procedure through which group reasoning is enacted, their participation is not agential in the sense required for this reasoning to be a plausible joint action; they are not the primary units of agency which direct this activity and which the activity itself serves. It is the rationality of the group itself which the activity seeks to reinforce and it is the agency of group itself which does the reinforcing. The contributions of individuals to this activity are not just subvenient but *subservient* to the rational agency of the group when it is instituted and sustained by means of reasoning. This means that when Pettit claims that group agents are not born but made through a process of group reasoning he introduces a puzzle about the genesis of group agents that he does not appear to have the resources to resolve.

3. Conclusion: Recognition and Group Agency

Pettit's account of the coming into being of group agents depicts this process as having two distinct stages. In the first stage we find that certain collectives, whose members share a common purpose, will eventually encounter situations of the sort exemplified by the Doctrinal Paradox – situations which pressurise them into taking steps to satisfy the demands of rationality at the collective level. This pressure can be seen as the realisation on the part of members of the collective that unless they make their collectively-held attitudes rational, they will find themselves unable to effectively pursue the purpose they all share. In the second stage, the members respond to this pressure by jointly acting to commit themselves to organisational mechanisms guaranteeing the rationality of their collective views; in this way, they make themselves into an autonomous group agent. I have sought to cast doubt especially upon the coherence of this second stage in the process: it does not seem plausible to me that individuals could jointly act so as to become, effectively, subpersonal machinery in a larger person.

If I am right then Pettit's account of the emergence of group agents is not compelling, or at least is not complete. In this final, concluding section I look beyond Pettit's own theory to see

whether some assistance can be brought in from elsewhere; my tentative suggestion is that this can be found in what Pettit would likely consider the most unwelcome of places.

3.1 A Hegelian Dilemma

The question we have been pursuing – ‘Who makes a group agent?’, or ‘Who is to collectivise reason?’ – is strikingly reminiscent of a question Hegel asks in the *Philosophy of Right* (1967: S273): ‘Who is to frame a constitution?’ The question seems to pose the following dilemma:

“either the would-be framers are an atomistic collection of individuals, who thus lack the moral unity that a political constitution presupposes, or they already identify themselves as a people, a *Volks*, in which case they already exist as a constituted “moral unity,” whose unwritten constitution exists in its *Volksgeist*” (Rehg 2011)

The first option makes the attempt to frame a constitution hopeless, while the second makes it unnecessary. Our question, ‘Who is to collectivise reason?’ can be seen to create a similar dilemma: either the would-be collectivisers are an atomistic heap of individuals who lack the unity that the process of collectivising reason requires (which makes the process hopeless), or they are already a unified group that can think and act as an integrated whole (which makes the process unnecessary). The various efforts made by Pettit and his collaborators to finesse the being-becoming circularity – especially the suggestion that joint action can account for the emergence of group agents – can thus be read as attempts to show up this dilemma as false. But, as we have seen, the jury (or the panel of judges, as the case may be) is still out on the success of these attempts.

How then does Hegel resolve the dilemma posed by his question? The dilemma is false, claims Hegel, since it presupposes that a constitution is something which can be *made*:

“it is absolutely essential that the constitution should not be regarded as something made, even though it has come into being in time. It must be treated rather as something simply existent in and by itself, as divine therefore, and constant, and so as exalted above the sphere of things that are made” (Hegel 1967, S273)

In short, the dilemma vanishes the moment we accept that ‘a people’ is not something which comes to be ‘constituted’ by their own or anyone else’s efforts; constitutions cannot be made

but can only be made explicit or, perhaps, amended. It is senseless to try to institute a constitution since the only possible subjects of a constitution are those who are already constituted as a people ('sharing a *Volksgeist*'), and hence already the holders of a constitution (though it may be implicit). But if constitutions cannot be instituted by human effort, then where do they come from? Hegel's suggestion is that they are 'divine' – that they are the product of what he elsewhere calls 'objective spirit' or *Geist*. In the case of the sort of collectives with which Pettit is primarily concerned, this would mean that they cannot bootstrap themselves into existence by instituting their own grouphood:

"we cannot be exercising an *instituting power* without taking ourselves to have been *instituted* as a "we." We cannot just institute ourselves into our instituting powers. Therefore, one has to acknowledge a primacy of... objective spirit over the subjective spirit" (Descombes 2011: 388)

3.2 A Hegelian Solution for Pettit

Is a similar move available to Pettit? If he were to give up on the claim that groups are 'made, not born' and acquiesce to some form of this Hegelian account of the genesis of group agents, would a route out *his* dilemma be forged? It certainly would, since it would allow him to say that the collectivisation of reason is something which simply happens – perhaps organically, perhaps through some or other divine means – but not something anybody (the collective, or its members) has to *achieve*. And once the group is up and running ('born'), the sorts of mechanisms we have been discussing can then intelligibly feature as means by which the group sustains itself.

But what is equally certain is that Pettit would emphatically reject this sort of solution. Indeed, one of what he considers to be the chief virtues of his account of group agency is that it can be sharply distinguished from the mysterious 'animation theory' about groups:

"the broadly Hegelian thought... [that] individual members give rise to a single agent only when a counterpart force comes on stream with its transforming effect" (List & Pettit 2011: 73-74)

Now I do not wish to consider the prospects for foisting 'animation theory' upon Pettit, though I do think that a different 'broadly Hegelian' approach might help us to diagnose the

source of Pettit's being-becoming trouble. The approach is inspired by Robert Brandom's (2002, 2009) social pragmatist reading of Hegel's master concept of 'mutual recognition', and what it helps to show is how the emergence of group agents might be a more *social* process than Pettit allows. Specifically, it seems plausible to think that the recognitive attitudes of others in a wider discursive community might have some role to play in the constitution of a group intentional subject.

We can begin with something which Pettit clearly accepts about the notion of a rational agent (or indeed a person): that it is a normative status. What this means is that being a rational agent (or person) implies that one is normatively *constrained* in certain ways and also that one is normatively *entitled* to adopt certain attitudes and take certain courses of action. But whom or what institutes these normative statuses, making them *genuine* or *binding*? According to Brandom, it is part of Kant's notion of the enlightenment ideal of autonomy that we realise that no such normative status would be genuine were it to be imposed from without (Brandom 2009: Ch 2).

The point can be illustrated by analogy with the chief development of enlightenment political philosophy – the idea of the social contract. Just as social contract theorists claimed that the legitimacy of political authority over citizens derives from those citizens' autonomous choice to subject *themselves* to that authority, so too do the norms which govern our thought and talk more generally only get a grip on us if they are somehow self-imposed. But this creates a puzzle, since it is hard to see how the norms of reason which rational creatures are subject to could be both self-imposed and genuinely binding: if we are the masters of our own boundedness then are we really *bound*? It is this puzzle – generated by Kant's dual insistence on normativity and autonomy – which Brandom sees Hegel as seeking to address via his notion of mutual recognition.

The solution wrought by Hegel involves the notion of a community:

"Hegel's principle innovation is his idea that in order to follow through on Kant's fundamental insight into the essentially *normative* character of mind, meaning, and rationality, we need to recognize that normative statuses such as authority and responsibility are at base *social* statuses" (Brandom 2009: 66)

The enlightenment ideal of autonomy entails that normative statuses are instituted by human attitudes, and for Brandom's Hegel, the key attitude is that of 'recognition' – the *treatment* of others as bearers of those statuses. What we need in order to actually be authoritative, or responsible, it to be recognised as such by others. But in seeking to be so recognised we are also recognising – we are according a certain kind of authority to those others we seek to be recognised by, namely, the authority to recognise us. According to Brandom's Hegel, normative statuses are only instituted when recognition is in this way symmetrical, or 'reciprocal'.

Reciprocal recognition thus provides not only a distinctively social account of what it is to *be* the bearer of certain normative statuses, it also provides a distinctively social account of what *becoming* the bearer of such statuses involves. Coming to be a good chess player, to borrow Brandom's own 'mundane' example, is a matter of coming to be recognised as such by those one recognises as such. Reciprocal recognition refers not just to an inter-subjective attitudinal *pattern* but to a certain sort of inter-subjective *process*:

"Achieving the status of being a good chess player is not something I can simply do by coming subjectively to adopt a certain attitude towards myself... It is up to me whom I recognise as good chess players, in the sense in which I aspire to be one. But it is not then in the same way up to me whether I qualify as one of them... To earn their recognition in turn I must be able to play up to their standards... My recognitive attitudes can define a virtual community, but only the reciprocal recognition by those I recognise can make me actually a member of it, accord me the status for which I have implicitly petitioned by recognising them" (Brandom 2009: 70-71)

Could something like this model of reciprocal recognition in principle be exploited by Pettit to overcome the problem of how group agents emerge? I think it could. For one thing it need not involve any of the metaphysical mysteriousness Pettit associates with Hegelianism: for Brandom's Hegel (2009: 72), *Geist* or 'spirit' is nothing other than the "realm of discursive activity" which is "synthesized by reciprocal recognition". Perhaps then we can see reciprocal recognition as achieved in part by the joint action of members in a collective: members might jointly petition a wider community for recognition as an intentional subject. If they succeed –

if the community defined by their joint act of petitioning reciprocates by recognising them as an intentional subject – then their status as such is secured. Pettit's individualistic 'performative' approach to personhood obscures this possibility, making it seem as if intentional subjects could only be made or born, and this is what lands him with the bootstrapping difficulty I think he struggles to overcome. But if there is something to be said for the role of recognitive attitudes in the social process of becoming a group agent, then perhaps we can conclude that group agents are neither born nor made, though it may take a village to raise one.

Afterword: Whither Group Agency?

In the introduction to this collection I characterised the aim of each essay in terms of a particular sort of question about group agency. Essay 1 was concerned with a double-edged *what-question* about the meaning of group agency: in what *sense* might suitably designed social groups be counted as intentional subjects or persons, and what are the conceptual *implications* of viewing them in this way? Essay 2 then shifted the focus from the ‘what’ to the ‘why’, inquiring after the *motivations* of group agency: *why* might some sets of individuals seek to transform themselves into group agents? Finally, Essay 3 addressed what I called the *how-question* about the mechanisms of group agency: just how can collectives go about transforming themselves in this way? In exploring what I take to be Pettit’s answers to these questions, I sought to cast doubt upon the notions of rationality and personhood implicit in his account, and the coherence of his description of the emergence of group agents. But I also tried throughout to supplement these negative efforts with something more positive, and this involved gesturing towards some Brandom-inspired communitarian solutions to each of the problems identified.

The methodological and normative individualism espoused by Pettit seems to create a number of difficulties for his account of group agents. For one, the insistence that a group agent is constituted by nothing other than the attitudes and activities of its individual members – essentially the supervenience thesis encountered in Essay 1 – effects what I dubbed a kind of ‘alienation’. Supervenience even of the ‘holistic’ sort entails that individual persons are cast in the role of homunculi or ‘sub-persons’ within a group agent, and this involves a radical recalibration of the way we think of ourselves. A second difficulty, discussed in Essay 2, relates to the rationality of group agents. In endorsing a conception of rationality which overlooks the ways in which the exercise of reason is an essentially *communal* affair, Pettit contends that groups can be subject to rational norms in just the same ways as individuals. But recognising, as Brandom suggests we should, that rationality is not a matter only of *responsibility* but also of *authority* means we can accept this while denying that it provides any motivation for group agency. Finally, as explored in Essay 3, Pettit’s account of the genesis of group agents suffers from a particular failing brought on by his allegiance to methodological individualism. Specifically, Pettit’s failure to appreciate how attitudes of

reciprocal recognition in a community might play a role in instituting the normative statuses of community members causes his account to lapse into a version of a familiar individualist myth – the myth of the self-made man.

Where should these considerations lead us? It should be clear that I think a less individualistic social ontology – one focused on communities rather than committees – is worth pursuing. Though I have remained agnostic on the issue of whether groups might be persons, I think that this issue depends on whether they could be *recognised* in a certain way rather than whether they are *realised* in a certain way, or whether they can themselves realise certain performances. In short, I feel that the concepts of personhood and rationality so central to the theory of group agency are not intelligible in the individualistic terms invoked by Pettit, but must instead be seen in communitarian terms. Construing them in this way need involve none of the metaphysical mysteriousness of ‘collectivism’, though it is bound to have significant social ontological consequences of its own. But what I have offered here is only (I hope) the prolegomenon to a fuller development of this non-mysterious communitarian alternative to Pettit’s theory.

Works Cited:

- Antonelli, G.A.** 2005. *Grounded Consequence for Defeasible Logic*, Cambridge: Cambridge University Press.
- Aristotle**, *De Anima*, in McKeon, R. (ed) 1941. *The Basics Works of Aristotle*. New York: Random House.
- Attneave, F.** 1960. 'In Defence of Homunculi', in W. Rosenblith (ed.), *Sensory Communication*. Cambridge: MIT Press, pp. 777-782.
- Bennett, M. & Hacker, P.** 2003. *Philosophical Foundations of Neuroscience*. London: Blackwell.
- Biro, J.** 1981. 'Persons as corporate entities and corporations as persons'. *Nature and System* 3 (September):173-80.
- Bratman, M.** 1999. *Faces of intention: Selected essays on intention and agency*. Cambridge: Cambridge University Press.
- 2009. 'Shared agency', in Mantzavinos, C. (ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*. Cambridge: Cambridge University Press.
- Brooks, D.** 1986. 'Group minds' *Australasian Journal of Philosophy* 64 (December):456-70.
- Block, N.** 1980. 'Troubles with Functionalism'. In Block, N. (ed) *Readings in Philosophy of Psychology*, Vol. I. Cambridge: Harvard University Press.
- 1995. 'On a Confusion About a Function of Consciousness'. *Brain and Behavioral Sciences* 18: 227-247
- Brandom, R.** 1994. *Making it Explicit*. Cambridge: Harvard University Press.
- 2002. *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge: Harvard University Press.
- 2009. *Reason in Philosophy: Animating Ideas*. Cambridge: Harvard University Press.
- 2012. 'A Hegelian Model of Legal Concept Determination'. Paper presented at 2012 Inland Northwest Philosophy Conference. Retrieved from: <http://www.webpages.uidaho.edu/inpc/15th-2012/>
- Chalmers, D.** 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chapman, B.** 2002. 'Rational Choice and Categorical Reason,' *Pennsylvania Law Review* 151: 1169-1210
- Clark, A.** 1994. 'Beliefs and Desires Incorporated'. *Journal of Philosophy* 91 (8): pp. 404-25.
- Crick, F.** 1995. *The Astonishing Hypothesis*. London: Touchstone Books.
- Davidson, D.** 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.
- Dennett, D.** 1969. *Content and Consciousness*. London: Routledge.
- 1978. *Brainstorms*. Boston: MIT Press.
- 1987. *The Intentional Stance*. Boston: MIT Press.
- 1991. *Consciousness Explained*. Penguin
- 1998. 'Self-Portrait'. In Brainchildren. Boston: MIT Press.
- 2007. 'Philosophy as Naive Anthropology: Comment on Bennett and Hacker'. in Bennett et al. *Neuroscience & Philosophy*. New York: Columbia University Press..
- 2009. 'Intentional Systems Theory', in McLaughlin et al (eds). *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press.

- Descombes, V.** 2011. 'The Problem of Collective Identity: The Instituting We and the Instituted We' in Laitinen, A & Ikaheimo, H. (eds). *Recognition and Social Ontology*. Leiden: Brill.
- Dworkin, R.** 1986. *Law's Empire*. Cambridge: Harvard University Press.
- Engel, P.** 2002. 'Intentionality, normativity, and community,' *Facta Philosophica* 4:25-49
- Fodor, J.** 1968. *Psychological Explanation*. New York: Random House.
- Frankish, K.** 2005. 'Non-Monotonic Inference' in K. Brown (ed.) *The Encyclopedia of Language and Linguistics*, 2nd Edition. Oxford: Elsevier.
- Garland, D.** 2001. *The Culture of Control*. Chicago: University of Chicago Press.
- Gilbert, Margaret.** 1989. *On Social Facts*. London: Routledge.
- 1996. *Living Together*. Maryland: Rowman & Littlefield.
- 2001. 'Collective Preferences, Obligations, and Rational Choice'. *Economics and Philosophy* 17: 109-120.
- Ginsberg, M.,** (ed.), 1987, *Readings in Nonmonotonic Reasoning*, Los Altos, CA: Morgan Kauffman.
- Goldman, A.** (2004) 'Group Knowledge Versus Group Rationality: Two Approaches to Social Epistemology,' *Episteme* 1(1): 11-22
- Hegel, G.** 1967. *Philosophy of Right*, trans. T. M. Knox, New York, Oxford University Press
- Hobbes, T.** 1993. *Leviathan*, Curley, E. (ed.), Indianapolis: Hackett.
- Holmes, O.** 1897. 'The Path of the Law', 10 Harv. L. Rev. 457, 461
- Hornsby, J.** 2000. 'Personal and Sub-Personal: A Defence of Dennett's Early Distinction.' *Philosophical Explorations* 3 (1):pp. 6-24.
- Huebner, B., Bruno, M. & Sarkissian, H.** 2010. 'What Does the Nation of China Think About Phenomenal States?' *Review of Philosophy and Psychology* 1 (2): pp.225-243.
- Kornhauser, L. & Sager, L.** 1986. 'Unpacking the Court,' *Yale Law Journal* 96: 82-117
- 1993. 'The one and the many: adjudication in collegial courts' *California Law Review*, 81:1-59
- 2004. 'Group Choice in Paradoxical Cases,' *Philosophy and Public Affairs* 32: 249-76
- Lewis, D.** 1983, 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61(4): 343-377.
- List, C.** 2005. 'Group knowledge and group rationality: a judgment aggregation perspective,' *Episteme* 2(1): 25-38.
- 2006. 'The Discursive Dilemma and Public Reason,' *Ethics* 116: 362-402
- List, C & Pettit, P.** 2002. 'Aggregating Sets of Judgements: An Impossibility Result,' *Economics and Philosophy* 18: 89-110
- 2005. 'On the many as one: A reply to Kornhauser and Sager'. *Philosophy and Public Affairs* 33 (4):377-390.
- 2006. 'Group Agency and Supervenience', in *Southern Journal of Philosophy* (Spindel Conference 2005), 44: 85-105
- 2011. *Group Agency*. Oxford: Oxford University Press.

- List, C & Puppe, C.** 2009. 'Judgment aggregation: a survey', in Anand, Pattanaik & Puppe (eds) *Handbook of rational and social choice*, Oxford: Oxford University Press
- Lycan, W.** 1979. 'A New Lilliputian Argument Against Machine Functionalism.' *Philosophical Studies* 35 (April): pp. 279-87.
- 1987. *Consciousness*. Boston: MIT Press
 - 1996. *Consciousness and Experience*. Boston: MIT Press.
- McDowell, J.** 1986. 'Functionalism and Anomalous Monism', in Le Pore & McLaughlin (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.
- McGeer, V.** 2002. 'Developing Trust,' *Philosophical Explorations* 5:1 (2002) 21-38
- Merton, R.** 1973. *The Sociology of Science*. Chicago: University of Chicago Press.
- Minsky, M.** 1985. *The Society of Mind*. New York: Touchstone Books.
- Nash, J. R.** 2003. 'A Context-Sensitive Voting Protocol Paradigm For Multimember Courts,' *Stanford Law Review* 56: 75-159
- Pettit, P.** 1993. *The Common Mind*. Oxford: Oxford University Press.
- 1995. 'The Cunning of Trust,' *Philosophy and Public Affairs*, 24: 202–225
 - 2001. 'Deliberative Democracy and the Discursive Dilemma,' in *Philosophical Issues* 11, pp. 268-99
 - 2002. 'Collective Persons and Powers', *Legal Theory*, 8 (2002), 443–470
 - 2003. 'Groups with Minds of Their Own,' *Socializing Metaphysics: the Nature of Social Reality*, Schmitt, F(ed.), Lanham, MD: Rowman & Littlefield: 167–93.
 - 2007. 'Rationality, Reasoning and Group Agency,' *Dialectica*, Vol 61, 2007, 495-519
 - 2009. 'The Reality of Group Agents,' in *Philosophy of the social sciences : philosophical theory and scientific practice*, C. Mantzavinos, ed., Cambridge, UK; New York : Cambridge University Press: 67–91.
- Pettit, P. & Schweikard, D.** 2006. 'Joint Actions and Group Agents,' *Philosophy of Social Sciences*, 36: 18–39
- Popper, K.** 1966. *The Open Society and Its Enemies*. London: Routledge & Kegan Paul.
- Quinton, A.** 1975. 'Social Objects,' *Proceedings of the Aristotelian Society* 75: 1–27
- Rehg, W.** 2011. Review of 'Recognition and Social Ontology'. *Notre Dame Philosophical Reviews*. Online: <http://ndpr.nd.edu/news/27905-recognition-and-social-ontology/>
- Rovane, C.** 1997. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton University Press.
- 2004. 'Alienation and the Alleged Separateness of Persons,' *The Monist* 87 (4):554-572.
- Russell, B.** 1919. *Introduction to Mathematical Philosophy*. Dover Publications.
- Ryle, G.** 1949. *The Concept of Mind*. Hutchinson and Co
- Scalia, A.** 1989. 'The Rule of Law as a Law of Rules', 56 U. Chi. L. Rev. 1175 (1989)
- Searle, J.** 1995. *The Construction of Social Reality*. New York, N.Y.: Free Press.
- Skinner, B.** 1971. *Beyond Freedom and Dignity*. Penguin Books.
- Tollefsen, D.** 2002a. 'Collective Intentionality and the Social Sciences'. *Philosophy of the Social Sciences* 32 (1): pp.25-50.
- 2002b. 'Organizations as True Believers'. *Journal of Social Philosophy* 33 (3): pp.395–410.

Tuomela, R. 1995. *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press.

- 2003. 'The We-Mode and the I-Mode,' in *Socializing Metaphysics: the Nature of Social Reality*, F. Schmitt, ed., Lanham, MD: Rowman & Littlefield: 93–127.

Wittgenstein, L. 1953. *Philosophical Investigations*. New York: Macmillan.